HUNGARIAN UNIVERSITY OF AGRICULTURE AND LIFE SCIENCES

PROJECT MANAGEMENT PROCESS MODEL FOR DIGITAL ERA BASED ON BIG DATA ANALYSIS

Doctoral (PhD) dissertation

Mehrzad AbdiKhalife

Gödöllő, Hungary

2024

Hungarian University of Agriculture and Life Sciences

Name of Doctoral School:

Doctoral School of Economic and Regional Sciences

Discipline:

Management and Business Administration

Head of Doctoral School:	**Prof. Dr. habil Zoltán Bujdosó, PhD**
	Full Professor
	Hungarian University of Agriculture and Life Sciences
Supervisor:	**Prof. Dr. habil Anna Dunay, PhD**
	Full Professor
	John von Neumann University

....	....
Approval of the Head of School	Approval of the Supervisor

TABLE OF CONTENTS

LIST OF FIGURES v
LIST OF TABLES vi
1. INTRODUCTION 1
1.1. Importance of the topic 1
1.2. Problem statement 3
1.3. Research questions/ research objectives 4
1.3.1. Research Questions 4
1.3.2. Research Objectives 5
1.4. Conceptual model / research model 5
2. LITERATURE REVIEW 7
2.1. General 7
2.2. Success in project management 9
2.3. Trends and challenges in project management 11
2.4. Project management methodologies and standards 13
2.5. Technology and project management 16
2.6. Bibliometric analysis 18
2.7. Data science 22
2.7.1. Fields of data science 23
2.7.2. Data science applications 24
2.7.3. Advantages and disadvantages of data science 26
2.7.4. Role and applications of data science in project management 28
2.8. Big Data 30
2.8.1. Big Data applications 31
2.8.2. Application, challenges, and trends of big data in project management 32
2.9. Natural Language Processing (NLP) 36
2.9.1. NLP applications 37
2.9.2. NLP applications in project management 38
2.9.3. Advantages of NLP applications in project management 40
2.9.4. Challenges of NLP applications in project management 41
2.10. Graph theory 42
2.10.1. Graph theory applications 44
2.10.2. Graph theory applications in project management 44
2.10.3. Challenges and limitations of graph theory in project management 45
2.11. Agile project management 47
2.11.1. Agile methodologies 48
2.11.2. Agile methodologies principles 49
2.11.3. Agile project management benefits 49
2.11.4. Agile project management challenges 51
2.11.5. Comparison of agile frameworks 52
2.11.6. Implementation of Agile Methodologies 53
3. MATERIALS AND METHODS 55
3.1. Data source 55
3.1.1. Academic source 55
3.1.2. Non-academic source 58
3.2. Application of graph theory 62
3.3. Using NLP in graph theory 67
3.4. Suggestion method to project manager 68
4. RESULTS AND THEIR DISCUSSION 71
4.1. Data collection 71
4.2. NLP implementation 76

4.3. Network creation 83
4.4. Network visualization 85
4.5. Network analysis 96
4.6. Thomas Kuhn science philosophy: 101
4.7. I-indicator C- indicator and ICCO ranking: 101
4.8. Relation between, Kuhn's philosophy and indicators: 102
4.9. Application of machine learning algorithms 106
4.10. Using machine learning algorithms as new metric 114
4.11. Validation 129
4.11.1. Random validation 130
4.11.2. Unseen Validation 132
5. CONCLUSIONS AND RECOMMENDATIONS 135
5.1. Conclusions 135
5.2. This study's contributions 137
5.3. Future expansion and recommendation: 137
6. NEW SCIENTIFIC RESULTS 143
7. SUMMARY 145
APPENDIX: BIBLIOGRAPHY 147

LIST OF FIGURES

1. Fig 1. The framework of data collection, preprocessing, feature mapping, and predictive analysis …………………… ……………………………………...…………………4
2. Fig 2. Conceptual model of the data analysis and prediction framework employed in this thesis …………………… ……………………………………………………….…6
3. Fig 3. Data collection tools and steps ………………………………………………….61
4. Fig 4. Web data collection from search to scrape steps ……………………………….62
5. Fig 5. The machine learning workflow: From collection to analysis ……………………69
6. Fig 6. A Preview of academic data ……………………………………………………...72
7. Fig 7. A Preview of non-academic data ………………………………………………74
8. Fig 8. Details of the generated networks for academic data ……………………………..84
9. Fig 9. Visualization of Academic data Keywords Unigrams ………………………… ...85
10. Fig 10. Visualization of Academic data Keywords Bigrams ……………………………86
11. Fig 11. Visualization of Academic data Keywords Trigrams ……………………………87
12. Fig 12. Visualization of Academic data Title Unigrams …………………………………88
13. Fig 13. Visualization of Academic data Title Bigrams …………………………………..89
14. Fig 14. Visualization of Academic data Title Trigrams …………………………………90
15. Fig 15. Visualization of Academic data Abstract Unigrams ……………………………91
16. Fig 16. Visualization of Academic data Abstract Bigrams ……………………………...92
17. Fig 17. Visualization of Non-Academic data Title Unigrams …………………………...93
18. Fig 18. Visualization of Non-Academic data Title Bigrams ……………………………94
19. Fig 19. Visualization of Non-Academic data Title Trigrams …………………………....95
20. Fig 20. Visualization of Non-Academic data Body Unigrams …………………………..96
21. Fig 21. Labelling procedure pseudo code …………………...……………………….119
22. Fig 22. Model accuracy and loss in imbalanced class data …………………………….123
23. Fig 23. Model accuracy and loss in balanced class data ………………………………123

LIST OF TABLES

1. Table 1. “Free” Node Ranking in consecutive quarters 118
2. Table 2. “Policy” Node Ranking in consecutive quarters....118
3. Table 3. “Engineering” Node Ranking in consecutive quarters....118
4. Table 4. “Background” Node Ranking in consecutive quarters 118
5. Table 5. “Productivity” Node Ranking in consecutive quarters 118
6. Table 6. “Include” Node Ranking in consecutive quarters 119
7. Table 7. MLP imbalanced Class Results.... 122
8. Table 8. MLP balanced Class Results.... 123
9. Table 9. SVM balanced Class Results 124
10. Table 10. SVM imbalanced Class Results 124
11. Table 11. Random Forest balanced Class Results.... 125
12. Table 12. Random Forest imbalanced Class Results 125
13. Table 13. XGBoost balanced Class Results 126
14. Table 14. XGBoost imbalanced Class Results.... 126
15. Table 15. The comparison between the predicted and actual class of the selected nodes (Class 1) 127
16. Table 16. The comparison between the predicted and actual class of the selected nodes (Class 0) 128
17. Table 17. The comparison between the predicted and actual class of the selected nodes (Class -1) 129
18. Table 18. Random Forest Validation Results (Random Validation) 131
19. Table 19. XGBoost Validation Results (Random Validation) 131
20. Table 20. XGBoost Validation Results (Unseen Validation) 133
21. Table 21. Random Forest Validation Results (Unseen Validation) 134

1. INTRODUCTION

1.1. Importance of the topic

In any scientific exploration, the foundational pillars that determine its course are the underlying needs and the ensuing motivations. The genesis of this research was not serendipitous; it was rooted in addressing distinct gaps and requirements observed in the domain. The needs that became evident emerged not just from theoretical voids, but also from practical challenges faced in real-world scenarios. On the other hand, the motivation to undertake this journey was fuelled by both the desire to contribute meaningful solutions to these challenges and a passion for advancing my understanding of the subject. In this discussion, I shall delve deeper into the intricacies of these needs and the driving forces behind the motivations that shaped the contours of this research.

Before embarking on my academic journey, my role as a practical project manager offered me a firsthand view of the complex landscape of project management. In this role, I frequently encountered the sheer volume of methods and techniques available to navigate the day-to-day challenges inherent in the profession. However, as is often the case with an abundance of choices, there arose a paradox. Instead of serving as an arsenal of tools aiding decision-making, the overwhelming array of information occasionally obscured the path, making crucial decisions more intricate than they needed to be. This realization became even more pronounced when considering that the essence of a project manager's role is to be an effective decision-making machine, consistently making informed choices amidst the cacophony of project variables. Transitioning into the realm of theoretical research for my Ph.D., I carried this insight with me, fueling my desire to dissect, understand, and potentially streamline the decision-making processes that are so central to effective project management.

In the intricate tapestry of project management, the role of a project manager stands out as a focal point, akin to a dynamic decision-making machine. Each decision made sets off a domino effect, impacting various facets of a project's lifecycle. To aid in this crucial task, a plethora of standards, software, and methodologies have been developed and published, each promising to streamline processes and enhance decision-making accuracy.

However, the blessing of these tools can also be a curse. The sheer volume and rapid evolution of available technologies can be overwhelming, leading to an overload of choices. This paradoxical situation places project managers in a challenging position: the need to continually update their knowledge and adapt to cutting-edge technologies while ensuring efficiency and top-tier performance in their current roles.

In essence, while tools and methodologies are indispensable, there's a pressing need to strike a balance. Project managers should harness the power of these resources without becoming ensnared in the complexity they sometimes introduce. As the adage goes, 'It's not about having the right tools, but about using the tools right.

On the flip side, the relentless march of technology presents its own set of challenges. Every day, I witness the emergence of novel methodologies, cutting-edge software, and innovative tools designed to redefine the paradigms of project management. This continuous influx of new information and techniques, while exciting, also brings with it an inherent pressure.

For project managers, staying abreast of the latest advancements isn't just about being informed; it's a requisite to ensure the continued relevance and efficacy of their strategies. However, the sheer pace of technological and methodological evolution can make this task daunting. It's akin to chasing a moving target: just as one becomes familiar with a new tool or approach, another emerges, demanding attention and understanding.

This rapid cycle of innovation and obsolescence not only requires project managers to be perpetual learners but also demands discernment. They must sift through the noise, determining which advancements truly hold value for their specific contexts and which might be distractions or temporary trends.

Embarking on a Ph.D. journey introduced me to the nuanced realities of academic research, a realm distinct from my prior experiences. Initially, I entered this pursuit fuelled by a vision to contribute significantly to my field. However, as I delved deeper into the academic environment, I recognized a pattern. The inception of a Ph.D. often begins with a broad vision. Yet, as years of rigorous study and research unfold within the university's walls, this vision undergoes refinement and crystallization, leading us, the Ph.D. candidates, to my definitive thesis topic.

However, an inherent challenge lurks within this process. By the time I meticulously craft and finalize my theses, the fast-paced evolution of knowledge and technology can render some of my findings less novel or even outdated upon graduation. This dynamic presents a perplexing facet of academic research: the potential for work to become antiquated almost as swiftly as it's conceived, emphasizing the transient nature of cutting-edge research.

Confronted with these multifaceted complexities, there was an evident need for a tool—one that could not only analyze the present landscape but also offer both theoretical researchers and practical practitioners a comprehensive map of the prevailing scenario. Such a tool would serve as a compass, guiding us through the intricate maze of evolving knowledge and methodologies.

In today's digital age, where the sheer volume of information can be both a boon and a challenge, specialized tools and methodologies are essential to navigate and comprehend the vast oceans of knowledge. Text mining stands out as a computational process that involves the extraction of valuable information from large volumes of text. By employing algorithms, it can discover patterns, themes, and relationships, revealing insights that might remain obscured when analyzing documents manually. Meanwhile, text analysis, which is more encompassing than text mining, uses natural language processing, computational linguistics, and other methodologies to understand structure, meaning, and intent in vast text datasets. This facilitates the discovery of trends and the extraction of specific details.

The classification of knowledge aids in segmenting and organizing the overwhelming amount of available information into coherent categories. By classifying knowledge, researchers can understand the structure of a domain, detect gaps, and identify overlaps. Parallel to this, monitoring the most accessed or discussed articles provides insights into the current trends and areas of heightened interest in the scientific community. These "hot topics" can guide researchers towards areas of pressing importance.

Scientometrics analysis is about measuring and analyzing scientific literature. It allows for the evaluation of the impact of publications and researchers and offers insights into the growth, development, and focus of different scientific fields. A related concept, bibliometrics, zeroes in on evaluating publication patterns, citation analyses, and the relationships between authors, institutions, and countries, offering insights into the influence and reach of scholarly works.

Finally, the network analysis of publications involves constructing and analyzing networks based on citations, collaborations, and co-authorships. By visualizing these networks, one can discern the relationships between researchers and institutions, illuminating the collaborative nature of science and pinpointing influential nodes in the network. All these tools, when employed judiciously, act as invaluable aids for researchers to navigate the intricate pathways of knowledge with precision and clarity.

While understanding the present landscape is invaluable, it only paints half the picture. Merely grasping the current state doesn't provide insights into the trajectory of future developments. Although predicting the future with precision remains elusive, recognizing patterns can offer glimpses into forthcoming trends. Consequently, my focus evolved, seeking to shed light on prospective pathways for both theoretical and practical project managers.

1.2. Problem statement

In the rapidly evolving digital era, the field of project management is confronted with a multitude of complexities. Agile project management, known for its flexibility and adaptability, has emerged as a popular approach to navigate this dynamic landscape. However, the integration of big data analysis into agile project management processes remains a complex and underexplored task. The need for a tool that can not only analyze the current landscape but also provide a comprehensive map of the scenario for both theoretical researchers and practical practitioners is evident.

This research aims to develop an Agile Project Management Process Model that effectively integrates big data analysis. This model is envisioned to serve as a compass, guiding us through the intricate maze of evolving knowledge and methodologies. It will help to navigate the vast oceans of knowledge, making sense of the sheer volume of information that characterizes the digital age.

The proposed model will incorporate text mining, a computational process that involves the extraction of valuable information from large volumes of text. By employing algorithms, text mining can discover patterns, themes, and relationships, revealing insights that might remain obscured when analyzing documents manually. The model will also encompass text analysis, which uses natural language processing, computational linguistics, and other methodologies to understand structure, meaning, and intent in vast text datasets. This facilitates the discovery of trends and the extraction of specific details.

The classification of knowledge, another key component of the proposed model, aids in segmenting and organizing the overwhelming amount of available information into coherent categories. By classifying knowledge, researchers can understand the structure of a domain, detect gaps, and identify overlaps. This process will be paralleled by monitoring the most accessed or discussed articles, providing insights into the current trends and areas of heightened interest in the scientific community. These “hot topics” can guide researchers towards areas of pressing importance.

Scientometrics analysis, which involves measuring and analysing scientific literature, will be integrated into the model. It allows for the evaluation of the impact of publications and researchers and offers insights into the growth, development, and focus of different scientific fields. A related concept, bibliometrics, zeroes in on evaluating publication patterns, citation analyses, and the relationships between authors, institutions, and countries, offering insights into the influence and reach of scholarly works.

Finally, the network analysis of publications will be incorporated into the model. This involves constructing and analysing networks based on citations, collaborations, and co-authorships. By visualizing these networks, one can discern the relationships between researchers and institutions, illuminating the collaborative nature of science and pinpointing influential nodes in the network.

The study will also investigate the impact of the proposed model on project outcomes, such as project completion time, cost, and quality. It will further explore the factors that influence the successful implementation of the model in different project environments.

The findings of this research could contribute to the field of project management by providing a comprehensive framework for effective decision-making amidst the cacophony of project variables. This could potentially enhance project outcomes and efficiency in the digital era. The framework of the research steps depicted in Fig. 1 offers a clearer perspective.

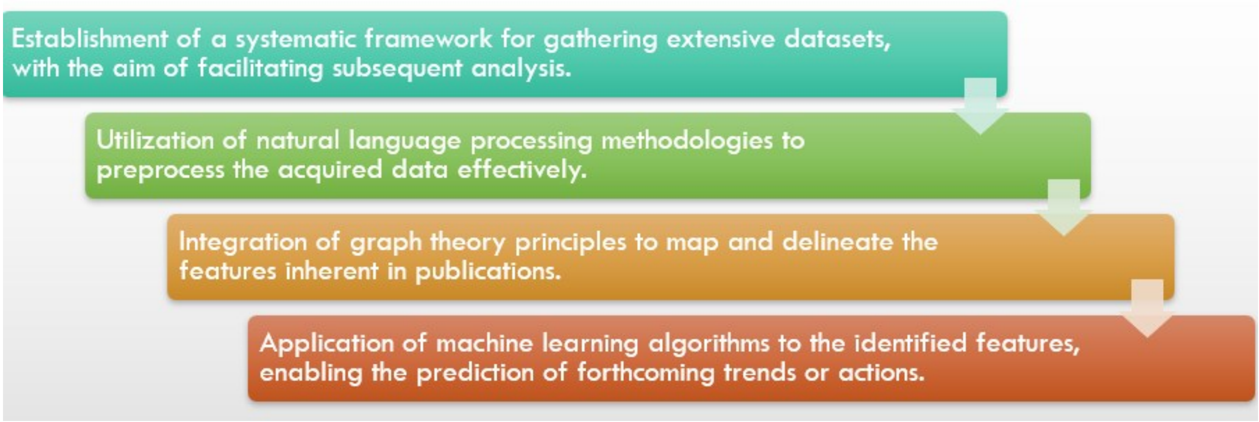


Figure 1. The framework of data collection, preprocessing, feature mapping, and predictive analysis (own research)

1.3. Research questions/ research objectives

1.3.1. Research Questions

1. De*coding and Anticipating Trends:* The first question that guides my research is: How can I leverage the combined power of network analysis and machine learning to decode and anticipate trends in word analysis? This question is not just about using these tools, but about exploring their synergistic potential. It's about understanding how these two methodologies can complement each other to provide a more comprehensive and accurate analysis of linguistic trends.
2. *Visualizing Knowledge:* The second question I aim to answer is: What insights can be derived from visualizing connections, patterns, and clusters in the current state of knowledge? This involves using network analysis to create visual representations of data, which can reveal patterns and connections that might otherwise remain hidden. It's about using these visualizations to identify prominent themes, spot emerging nodes of importance, and understand the relationships that shape the discourse.
3. *Data Extraction and Analysis:* The third question guiding my research is: How can web scraping techniques be used to extract articles from openly accessible websites, and what does this data reveal about global discourse and scholarly communication? This question is about the process of data collection and the insights that can be gleaned from the data. It involves using web scraping techniques to gather a comprehensive dataset and then analyzing this data to understand the breadth and depth of global discourse.
4. *Temporal Analysis:* The fourth question I seek to answer is: What shifts in discourse, emerging topics, and fading trends can be identified by comparing articles published in different quarters? This question is about understanding the dynamic nature of discourse and how it evolves over time. It involves segmenting the data based on publication quarters and tracking the shifts in discourse over these periods.
5. *Predictive Analysis:* The final question that guides my research is: How can a temporal lens, when combined with predictive machine learning models, help me understand the present narrative landscape and make informed projections about its future evolution? This question is about forecasting. It's about using the insights gained from the temporal analysis and the predictive power of machine learning models to make informed projections about the future evolution of the narrative landscape.

1.3.2. Research Objectives

1. *Harnessing Analytical Tools:* My first objective is to harness the synergistic power of network analysis and machine learning prediction models for decoding and anticipating trends in word analysis. This involves exploring the capabilities of these tools, understanding how they can complement each other, and using them to provide a comprehensive and accurate analysis of linguistic trends.
2. *Understanding the Knowledge Landscape:* My second objective is to use network analysis as a tool for mapping out and understanding the intricate landscape of the current state of knowledge. This involves creating visual representations of data, identifying prominent themes, spotting emerging nodes of importance, and understanding the relationships that shape the discourse.
3. *Data Sourcing and Extraction:* My third objective is to source data directly from the World Wide Web and employ web scraping techniques to extract articles from openly accessible websites. This involves gathering a comprehensive dataset and analyzing this data to understand the breadth and depth of global discourse.
4. *Temporal Segmentation:* My fourth objective is to segment data based on publication quarters for a more granular and temporal perspective. This involves tracking the shifts in discourse over these periods and understanding how the discourse evolves over time.
5. *Tracking Discourse Shifts:* My fifth objective is to track shifts in discourse, detect emerging topics, and identify fading trends by comparing articles published in different quarters. This involves using the insights gained from the temporal analysis to understand the dynamic nature of discourse.
6. *Predictive Analysis:* My final objective is to combine a temporal lens with predictive machine learning models for understanding the present narrative landscape and making informed projections about its future evolution. This involves using the predictive power of machine learning models to make informed projections about the future evolution of the narrative landscape.

By addressing these research questions and objectives, I aim to contribute significantly to the field of discourse analysis. My research will provide a comprehensive understanding of the current narrative landscape and offer valuable insights into its future evolution. I hope that my findings will serve as a valuable resource for scholars, policymakers, and practitioners interested in global discourse and scholarly communication. I am excited about the potential of this research to advance my understanding of global discourse and look forward to sharing my findings with the broader academic community.

1.4. Conceptual model / research model

This research proposes an innovative and dynamic conceptual model that seamlessly integrates big data analysis into the Agile Project Management Process. The model is designed with the aim to harness the transformative power of big data to enhance decision-making processes and improve project outcomes in agile project management.

The model will utilize network analysis as a primary tool to map out and comprehend the current state of knowledge in the field. By visualizing connections, patterns, and clusters, the model can discern relationships, identify prominent themes, and detect emerging nodes of importance.

The data for the analysis will be sourced directly from the expansive reservoir of the World Wide Web. Sophisticated web scraping techniques will be employed to extract articles from openly accessible websites. This approach will provide a comprehensive dataset that reflects global discourse and scholarly communication in the field of project management.

To provide a more granular perspective and to track temporal changes in discourse, the data will be segmented based on publication quarters. This segmentation will allow us to compare articles published in different quarters, enabling us to track shifts in discourse, detect emerging topics, and identify fading trends. This temporal lens will provide a dynamic view of the evolving landscape of knowledge.

The model will also incorporate machine learning prediction models. These models will be trained on the segmented data, learning from the patterns and trends in past and present discourse to predict future trends. This predictive capability will enable the model to anticipate changes in the knowledge landscape and provide timely insights into future trends in project management.

The findings of this research could contribute to the field of project management by providing a comprehensive framework for integrating big data analysis into agile project management processes. This could potentially enhance project outcomes and efficiency in the digital era. The framework of the research steps, as illustrated in Fig. 2, provides a clearer understanding of the overall process.

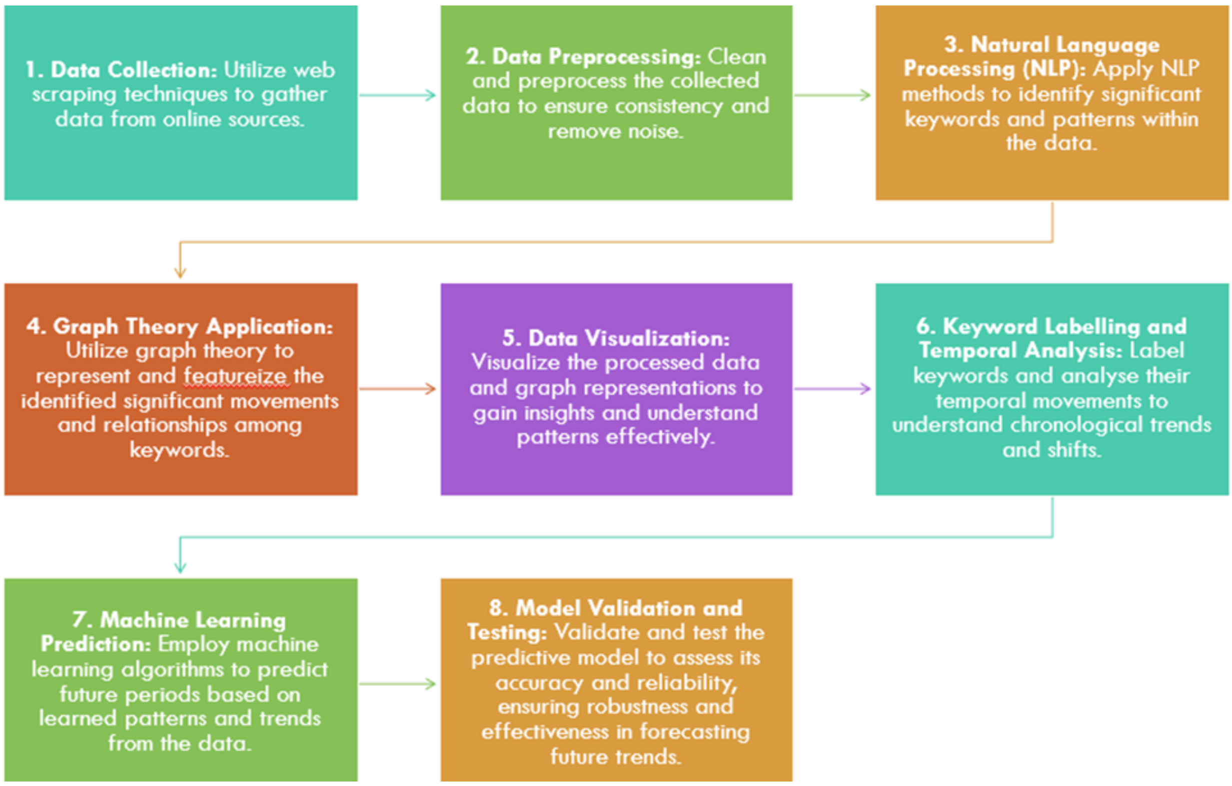


Figure 2. Conceptual model of the data analysis and prediction framework employed in this thesis (own research)

2. LITERATURE REVIEW

2.1. General

Project management is not a new idea and has been around for thousands of years. It is a critical practice that has been used throughout human history, with examples dating back to the construction of the ancient wonders of the world. Over the centuries, project management has evolved into a formal discipline with its own set of tools, methodologies, and best practices. The earliest known examples of project management can be found in the construction of the ancient wonders of the world. These massive undertakings required extensive planning, coordination, and the use of advanced engineering techniques. The Great Pyramid of Giza, which was built in 2570 BC and is considered a remarkable achievement of ancient project managers (Breyter 2022; Xue et al. 2022).

Project management involves planning, organizing, and executing projects to achieve specific goals within a set timeframe and budget. It requires the use of sophisticated mathematical calculations and various tools and techniques to keep track of progress, manage resources, and ensure, that tasks are completed efficiently. Building such a massive (Great Pyramid of Giza) structure required meticulous planning and organization, with workers having to move huge blocks of stone over long distances and up steep inclines (Ghorbani & Professor 2023).

The project managers of the time must have had exceptional skills in logistics, resource management, and leadership to oversee such a complex project successfully. They would have needed to coordinate the efforts of thousands of workers and ensure that everything was done according to plan. The example of the Great Pyramid of Giza illustrates that project management is a timeless concept that has been utilized throughout human history. While the tools and techniques may have evolved over time, the fundamental principles remain the same (Ahmed, Philbin & Cheema 2021; Ghorbani & Professor 2023).

While the historical evolution of project management highlights its timeless relevance, it also underscores the importance of continual adaptation and innovation in the field. In my opinion, modern project management, although rooted in ancient practices, has significantly transformed to meet the demands of an increasingly complex and fast-paced world. Today, project managers not only rely on traditional skills such as planning and coordination but also leverage cutting-edge technologies and methodologies to enhance efficiency and effectiveness. The integration of digital tools, data analytics, and agile frameworks has revolutionized how projects are managed, allowing for greater flexibility and responsiveness to change. This evolution reflects a broader trend in various disciplines where historical knowledge provides a foundation, but continuous improvement and adaptation drive progress. It is this dynamic interplay between tradition and innovation that ensures project management remains a critical practice for achieving success in diverse and evolving environments.

In the modern era, project management has become a critical component of many industries, including construction, manufacturing, aerospace, and technology. The practice has evolved significantly over the past century, with the development of new tools, methodologies, and best practices. One of the early pioneers of modern project management was American engineer Henry Gantt, who developed the Gantt chart in the early 20th century. The Gantt chart provided a visual representation of a project's schedule, helping managers to plan and coordinate complex projects more effectively (Seymour & Hussein 2014; Wilson 2003).

During World War II, the need for efficient project management became even more apparent as the military was forced to oversee massive projects like the Manhattan Project, which developed the atomic bomb. The military began to use more advanced project management techniques like critical path analysis, which helped them coordinate complex projects with tight deadlines. In the post-war era, project management began to spread beyond the military and into the private sector.

The 1950s and 1960s saw the development of the Program Evaluation and Review Technique (PERT) and the Critical Path Method (CPM), which became widely used in industries like construction, aerospace, and manufacturing (Forcael et al. 2014).

The 1980s and 1990s saw the emergence of project management as a distinct profession, with the creation of organizations like the Project Management Institute (PMI) and the development of widely recognized certifications like the Project Management Professional (PMP) credential. During this time, project management also began to incorporate more advanced tools like computer-aided design (CAD) and project management software.

In my opinion, the evolution of project management into a critical component of various industries highlights its growing significance in our interconnected world. The progression from early tools like the Gantt chart to advanced project management software and methodologies such as PERT and CPM reflects a trend towards increasing efficiency and sophistication. Modern project managers must adeptly navigate complex, globalized environments, leveraging digital platforms and technologies like AI and big data analytics. The professionalization of the field, marked by organizations like PMI and certifications like PMP, ensures standardized best practices and high competency, driving innovation and success across sectors. This continuous advancement underscores the enduring relevance and adaptability of project management in addressing contemporary challenges and opportunities.

In the 21st century, project management has continued to evolve as technology has advanced and businesses have become more global and complex. Agile project management methodologies like Scrum and Kanban have emerged to help manage fast-moving projects with changing requirements, while virtual project management tools have made it possible to coordinate teams across time zones and geographies. The history of project management is a long and varied one that has evolved over thousands of years. From the construction of the ancient wonders of the world to the development of advanced project management software, the practice of project management has played a vital role in the success of some of the most impressive projects in human history (Ahmed et al. 2020; Fan et al. 2022).

ISO 21500, developed by the International Organization for Standardization (ISO), provides guidance on the fundamentals of project management, including project initiation, planning, execution, monitoring and control, and project closure. It is designed to be used in conjunction with other ISO standards related to project management, such as ISO 10006 (quality management) and ISO 31000 (risk management). Both of these standards provide a framework for managing projects effectively and efficiently, and can help organizations improve project outcomes, reduce costs, and increase stakeholder satisfaction (Takagi & Varajão 2021).

The PMBoK and ISO 21500 are not the only project management standards available to practitioners. Other standards include Prince2 (developed in the UK), AgilePM (developed by the Agile Business Consortium), and the Scrum framework (developed by Jeff Sutherland and Ken Schwaber). Prince2 is a project management methodology that emphasizes controlled processes, clear roles and responsibilities, and regular progress reviews (Martin 2023; Timinger, Schmidtner & Reiche 2022). AgilePM is an agile project management framework that emphasizes collaboration, flexibility, and rapid iteration. Scrum is another agile framework that focuses on iterative development, daily stand-up meetings, and cross-functional teams (Chandrachooodan, Radhika & Palappan 2023; Herath & Perera 2023).

While these standards differ in their approach and methodology, they all share a common goal of improving project outcomes and delivering value to stakeholders. Effective project management requires a range of skills and competencies, including leadership, communication, problem-solving, and decision-making. They must also be able to manage resources effectively, including budgets, personnel, and materials (Britto et al. 2018).

In my opinion, the continuous evolution of project management methodologies and standards underscores the field's critical role in navigating the complexities of the modern business landscape. Agile methodologies like Scrum and Kanban, along with virtual project management tools, have revolutionized how teams adapt to rapid changes and coordinate across global time zones. Standards such as ISO 21500, PMBoK, Prince2, AgilePM, and Scrum offer structured frameworks that enhance project outcomes, reduce costs, and increase stakeholder satisfaction. Despite their different approaches, these methodologies share the common goal of delivering value and improving project success. Effective project management today demands a blend of leadership, communication, problem-solving, and resource management skills, highlighting the profession's multifaceted nature and its importance in driving innovation and efficiency in diverse industries.

2.2. Success in project management

There are many different factors that can influence the success of a project. Some of the most important of them are as follows:

Risk management is one of the critical aspects of project management. Project managers must be able to identify potential risks early on and develop strategies to mitigate them. They must also be able to adapt quickly to unexpected challenges and changes in project scope or requirements. Effective monitoring and evaluation are essential for successful project management. By tracking project progress and performance, project managers can identify potential issues and make adjustments as necessary to keep projects on track and on budget (Ashkanani & Franzoi 2022a; Fobiri, Musonda & Muleya 2022).

Effective stakeholder engagement is another important aspect of project management. Project managers must be able to communicate effectively with stakeholders throughout the project lifecycle, from initiation to closure. This includes identifying stakeholders, developing communication plans, and managing stakeholder expectations. Project managers must also be able to manage remote teams effectively, given the increasing trend towards remote work in many industries. This requires the use of advanced collaboration tools and techniques to keep teams connected and productive, as well as strong leadership skills to motivate and inspire team members. Moreover, there has been a growing emphasis on emotional intelligence in project management. Project managers who possess high levels of emotional intelligence are better able to manage team dynamics, communicate effectively with stakeholders, and resolve conflicts when they arise (Larson & DeChurch 2020).

In my opinion, the success of a project is influenced by a multitude of factors, each contributing to the overall effectiveness and efficiency of project management. Risk management is paramount, requiring project managers to anticipate potential challenges and devise proactive strategies. Adaptability is also crucial, as unexpected changes in project scope or requirements necessitate swift and decisive action. Monitoring and evaluation ensure that projects remain on track and within budget, enabling timely identification and resolution of issues. Effective stakeholder engagement, facilitated by clear communication and expectation management, is essential throughout the project lifecycle. The rise of remote work demands advanced collaboration tools and techniques, alongside strong leadership to maintain team cohesion and productivity. Additionally, emotional intelligence has become increasingly important, with project managers who exhibit high emotional intelligence being better equipped to manage team dynamics, communicate with stakeholders, and resolve conflicts. These diverse elements collectively underscore the multifaceted nature of successful project management in today's complex and dynamic business environment.

Another important consideration in project management is effective resource allocation, including budgeting, staffing, and equipment procurement. Effective resource management can help ensure that projects are completed on time and within budget, while also reducing waste and maximizing efficiency. Furthermore, project managers must be able to balance the needs of different stakeholders and manage competing priorities effectively. This requires strong negotiation skills and the ability to communicate effectively with a wide range of stakeholders (Berrone et al. 2023; Govindaras et al. 2023).

Effective change management is also critical for successful project management. Project managers must be able to anticipate and manage changes to project scope, schedules, and budgets, while also ensuring that these changes align with broader organizational goals and objectives. Effective governance and portfolio management can help organizations prioritize and allocate resources more effectively, while also reducing risk and maximizing value. Furthermore, there has been a growing emphasis on social responsibility and sustainability in project management. By incorporating sustainable practices into project planning and execution, project managers can help promote environmental stewardship and social equity while also driving business success (Rezvani et al. 2023; Shaikh & Randhawa 2022).

Project management is a complex process that involves the coordination of resources, time, and budget to achieve specific objectives. Over the years, the field of project management has grown significantly, and project management processes have evolved to include best practices and methodologies that enable organizations to achieve successful project outcomes. The project management process can be defined as a series of interconnected stages that are undertaken to achieve a specific goal within a defined timeframe. Each stage of the process is critical to the success of the project, and effective project management requires careful planning, coordination, and communication throughout the project lifecycle (Viti et al. 2022; Zhang et al. 2023).

In my opinion, effective resource allocation is a cornerstone of successful project management, encompassing budgeting, staffing, and equipment procurement. Proper resource management not only ensures timely and cost-efficient project completion but also reduces waste and maximizes efficiency. Balancing the needs of various stakeholders and managing competing priorities demand strong negotiation and communication skills. Change management is equally crucial, requiring project managers to handle adjustments in scope, schedules, and budgets while aligning these changes with organizational goals. Governance and portfolio management help in prioritizing and allocating resources effectively, mitigating risks, and enhancing value. The increasing focus on social responsibility and sustainability in project management is vital, as incorporating sustainable practices can promote environmental stewardship and social equity, alongside driving business success. Overall, project management is a multifaceted process that requires meticulous planning, coordination, and communication throughout the project lifecycle, reflecting its evolution into a discipline enriched with best practices and methodologies aimed at achieving successful project outcomes.

The project management process involves five stages, including initiation, planning, execution, monitoring and controlling, and closure. Each stage of the process requires specific tasks to be completed, and the successful completion of each stage is critical to the success of the project (Leong et al. 2023).

The initiation stage is the first phase of the project management process, where the project's feasibility is assessed, and the project's objectives are determined. During the initiation stage, project managers must ensure that they have a clear understanding of the project's goals, objectives, and scope. They must also identify the project stakeholders and ensure that they are aligned with the project's objectives (Amirtash, Parchami Jalal & Jelodar 2021).

The planning stage is the second phase of the project management process, where the project's scope is defined, and the project plan is developed. The project plan provides a detailed roadmap for the project's execution, including resource allocation, budgeting, scheduling, and risk management. During the planning stage, project managers must ensure that they have a clear understanding of the project's objectives, scope, and requirements. This requires the involvement of all stakeholders, including the project team, sponsors, and end-users (AlMaazmi 2022; Ikudayisi et al. 2022).

The execution stage is the third phase of the project management process, where the project plan is put into action. This stage involves the coordination of resources and activities to achieve the project's objectives. During the execution stage, project managers must ensure that they have adequate resources to complete the project on time and within budget. They must also ensure that the project team is aligned with the project plan and that they have the necessary skills and experience to complete their tasks (Liao, Yang & Quan 2023; Pan et al. 2023).

In my opinion, the structured framework of the project management process, comprising initiation, planning, execution, monitoring and controlling, and closure, forms a critical foundation for project success. The initiation phase, by assessing feasibility and defining objectives, sets a clear direction for the project and aligns stakeholders towards common goals. During planning, the detailed project plan serves as a roadmap, guiding resource allocation, budgeting, and risk management efforts, ensuring all aspects are meticulously addressed before execution begins. Execution, the phase where plans are put into action, underscores the importance of resource coordination and team alignment to achieve project milestones effectively and within set timelines and budgets. Each stage's seamless integration and effective management not only facilitate project progress but also enhance organizational adaptability and stakeholder satisfaction, demonstrating the essential role of structured project management methodologies in driving successful project outcomes.

The monitoring and controlling stage are the fourth phase of the project management process, where the project's progress is tracked, and any necessary adjustments are made. This stage involves regular performance evaluations to ensure that the project stays on track and that any issues or risks are addressed. During the monitoring and controlling stage, project managers must closely monitor the project's progress and make any necessary adjustments to the project plan (Amir, khodeir & Khaled 2023; Romzi & Ing 2022).

The closure stage is the final phase of the project management process, where the project's success is evaluated, and the project is officially completed. This stage involves conducting a project review to determine what worked well and what could be improved in future projects. The closure stage provides an opportunity to learn from the project and to ensure that best practices are implemented in future projects (Bian & Tolford 2022; Guanci & Bjork 2019).

In my opinion, the monitoring and controlling stage of the project management process is essential for tracking progress and making necessary adjustments to ensure project alignment with goals and effective risk management. The closure stage provides a critical opportunity to evaluate project success, identify lessons learned, and implement best practices for future projects, fostering continual improvement and organizational growth. These final phases not only mark project completion but also contribute significantly to enhancing project outcomes and overall operational excellence.

2.3. Trends and challenges in project management

One of the significant trends in project management is the move towards hybrid project management models that incorporate elements of both traditional and agile methodologies. This allows organizations to take advantage of the benefits of both approaches while minimizing their respective drawbacks. The rise of remote work has also led to an increased focus on virtual collaboration and remote project management. As more teams work from home or other locations, project managers must utilize a variety of tools and techniques to keep their teams aligned and productive. Furthermore, the ongoing COVID-19 pandemic has had a significant impact on project management, with many organizations forced to adapt quickly to new working conditions and changing project requirements. To address these challenges, project managers are increasingly turning to innovative solutions such as advanced collaboration software, automation, and artificial intelligence. These technologies can help streamline project management processes and allow project managers to focus on higher-level tasks that require human expertise (Abioye et al. 2021; Leyer & Schneider 2021).

Furthermore, project managers are also recognizing the importance of soft skills such as communication, leadership, and conflict resolution. While technical skills remain essential, it is increasingly recognized that successful project management requires a holistic approach that incorporates both technical and interpersonal skills (Khatib, Almteiri & Qasemi 2021).

As businesses become increasingly complex and global, project management will continue to play a vital role in ensuring successful outcomes for projects of all sizes and scopes. There is a growing recognition of the importance of project management in driving innovation and growth. By adopting new technologies, processes, and practices, project managers can help their organizations stay ahead of the curve and achieve long-term success (Gemino, Horner Reich & Serrador 2020; Vrchota et al. 2020).

In my view, a significant trend in project management is the adoption of hybrid models that blend traditional and agile methodologies, offering flexibility and efficiency while minimizing drawbacks. The shift to remote work has emphasized the need for virtual collaboration tools and strategies to maintain team productivity across diverse locations. The COVID-19 pandemic has accelerated the use of advanced technologies like collaboration software and AI, enabling streamlined project management and freeing up time for strategic decision-making. Soft skills such as communication and leadership are increasingly recognized as crucial alongside technical expertise, ensuring holistic project success. Looking forward, project management's role in driving innovation and organizational success through adaptable practices and technologies remains pivotal in today's complex and global business landscape.

Overall, project management is a dynamic and rapidly evolving field that is essential to the success of projects across all industries. The emergence of new technologies, methodologies, and work practices is transforming the way projects are planned, executed, and managed. Project managers who can adapt to these changes and develop the necessary skills and expertise will be well-positioned to thrive in this exciting and challenging field. However, despite these challenges, project management has become increasingly important as organizations seek to innovate, adapt, and succeed in today's rapidly changing business environment. Project management helps ensure that projects are completed on time, within budget, and to the satisfaction of stakeholders (Kermanshachi, Nipa & Dao 2021; Zayat & Senvar 2020).

Furthermore, project managers must be able to adapt quickly to changing circumstances and unforeseen challenges. This requires a high degree of flexibility and agility, as well as the ability to think creatively and come up with innovative solutions to complex problems. Moreover, effective leadership is essential for successful project management. Project managers must be able to motivate and inspire their teams, provide direction and guidance, and make difficult decisions when necessary. This includes being able to manage team dynamics effectively and build a culture of collaboration, trust, and accountability (Hamouche 2021).

In my view, project management is pivotal for success across industries, constantly evolving with new technologies and methodologies. Adaptability and skill development are crucial for thriving in this dynamic field. Despite challenges, project management's role in driving innovation and delivering on stakeholder expectations is paramount. Flexibility, agility, and creative problem-solving are essential qualities, alongside effective leadership that inspires teams and fosters collaboration. Project management remains indispensable for organizations aiming to achieve strategic goals and navigate today's competitive business environment.

Another important consideration in project management is the use of technology. By leveraging tools and software such as project management software, online collaboration platforms, and data analytics, project managers can streamline processes, improve efficiency, and increase the likelihood of project success. In addition, project managers must be able to balance the needs of multiple stakeholders, including clients, team members, and other parties. This requires strong interpersonal skills and the ability to negotiate and manage conflicting priorities effectively. Moreover, project managers must be able to plan and execute projects in a way that maximizes value and minimizes risk. This requires a deep understanding of business strategy and a willingness to innovate and experiment with new approaches and technologies. Another important aspect of project management is ongoing monitoring and evaluation. By tracking project progress, identifying potential issues, and making adjustments as necessary, project managers can ensure that projects remain on track and meet their intended outcomes (Elia, Margherita & Secundo 2020; Stone & Stone 2023).

Furthermore, project managers must be able to adapt to changing market conditions and emerging trends in order to stay ahead of the curve and drive innovation within their organizations. Finally, ongoing professional development and training are essential for successful project management. As new technologies and methodologies emerge, project managers must stay up-to-date with the latest trends and tools in order to remain effective and competitive. Overall, project management is a critical function that plays a key role in driving the success of projects across all industries. The ongoing evolution of project management is driven by changes in technology, work practices, and business needs. Project managers who can adapt to these changes and develop the necessary skills and expertise will be well-positioned to thrive in this exciting and dynamic field (Alfreahat & Sebestyén 2022; Martin 2023).

In my view, technology is integral to modern project management, facilitating efficiency and success through tools like project management software and online collaboration platforms. Effective project managers must navigate diverse stakeholder needs with strong interpersonal skills and negotiation abilities. Strategic planning, innovation, and risk management are essential for maximizing project value while minimizing risks. Ongoing monitoring allows for adjustments to keep projects aligned with desired outcomes. Adapting to market changes and continuous professional development are crucial for staying competitive in the evolving landscape of project management across industries.

2.4. Project management methodologies and standards

Existence of the different methodologies is necessary to have a discipline for project management. These documents outline essential methodologies for project managers.

The PERT/Cost system document, which was published by the DOD and NASA in 1962, marked a significant milestone in the development of project management as a discipline. This document outlined a methodology for planning and controlling complex projects that became widely adopted across both the public and private sectors. The PERT system was designed to help manage the complexities of large-scale projects, such as those involved in space exploration and defense programs. It involves breaking down a project into smaller, more manageable tasks and then estimating the time required to complete each task. By identifying the critical path of a project, project managers could determine the sequence of tasks that must be completed on time in order to meet project deadlines (Alvesson & Deetz 2011; Alvesson & Willmott 2012).

The critical path method, which was also developed in the 1960s by Kelley and Walker, is another well-known project management technique. The critical path method involves identifying the optimal sequence of activities required to complete a project on time and within budget, and then optimizing resource allocation and scheduling to ensure that these activities are completed efficiently. The integration of WBS, PERT, and CPM into project management standards helped establish a common language and methodology for managing projects across different industries and sectors. This standardization helped improve project outcomes by providing a consistent framework for project planning, execution, and control (Essam, Khodeir & Fathy 2023; Zhao et al. 2023).

In addition to these established methodologies, there has been ongoing innovation and experimentation in the field of project management in recent years. Agile methodology, for example, has emerged as a popular approach to project management in industries such as software development, where rapid iteration and collaboration are essential for success. Other emerging trends in project management include the use of artificial intelligence and machine learning tools to automate routine tasks and improve decision-making, as well as an increased focus on sustainability and social responsibility in project planning and execution (Leyer & Schneider 2021).

In my opinion, the existence of various methodologies is crucial for establishing project management as a disciplined practice. Documents like the PERT/Cost system, originating from collaborations between the DOD and NASA in 1962, marked significant milestones by providing structured approaches to plan and control complex projects. These methodologies, including PERT and the critical path method developed by Kelley and Walker, offer frameworks to manage project complexities by breaking them into manageable tasks and optimizing resource allocation and scheduling. Standardizing these methodologies through integration into project management practices has enhanced project outcomes by ensuring consistency in planning, execution, and control across diverse industries. Moreover, ongoing innovations such as Agile methodology and the integration of artificial intelligence reflect evolving trends aimed at improving project efficiency and addressing contemporary challenges like sustainability and social responsibility.

Moreover, there has been a growing emphasis on project governance and portfolio management, which involve establishing clear roles and responsibilities for project stakeholders and ensuring that projects align with broader organizational objectives. Effective governance and portfolio management can help organizations prioritize and allocate resources more effectively, while also reducing risk and maximizing value. Effective communication remains at the heart of successful project management. Project managers must be able to communicate clearly and effectively with team members, clients, vendors, and other stakeholders in order to ensure that everyone is aligned on project goals, timelines, and expectations. This requires strong interpersonal skills, as well as an understanding of different cultural norms and working styles.

In addition, project managers must be able to manage remote teams effectively, given the increasing trend towards remote work in many industries. This requires the use of advanced collaboration tools and techniques to keep teams connected and productive, as well as strong leadership skills to motivate and inspire team members. Furthermore, ongoing monitoring and evaluation are essential for successful project management. By tracking project progress, identifying potential issues, and making adjustments as necessary, project managers can ensure that projects remain on track and meet their intended outcomes (Fobiri, Musonda & Muleya 2022; Stone & Stone 2023). This requires a continuous improvement mindset, as well as a willingness to experiment and try new approaches.

Moreover, there has been a growing emphasis on emotional intelligence in project management. Project managers who possess high levels of emotional intelligence are better able to manage team dynamics, communicate effectively with stakeholders, and resolve conflicts when they arise. Another important consideration in project management is effective resource allocation, including budgeting, staffing, and equipment procurement. Effective resource management can help ensure that projects are completed on time and within budget, while also reducing waste and maximizing efficiency (Alzoubi & Aziz 2021).

In my view, project governance and portfolio management are increasingly vital for aligning projects with organizational objectives and optimizing resource use. Effective communication remains essential for ensuring clarity among stakeholders, while managing remote teams requires robust leadership and collaboration tools. Continuous monitoring and adaptive management are crucial for staying on course amid evolving project needs. Emotional intelligence enhances team dynamics and conflict resolution, while efficient resource allocation supports timely and cost-effective project delivery.

Furthermore, project managers must be able to balance the needs of different stakeholders and manage competing priorities effectively. This requires strong negotiation skills and the ability to communicate effectively with a wide range of stakeholders. Effective risk management is also critical for successful project management. Project managers must be able to identify potential risks early on and develop contingency plans to mitigate them, while also being able to adapt quickly to unexpected challenges (Fobiri, Musonda & Muleya 2022; Najib et al. 2022).

While the PMBoK, ISO 21500, Prince2, AgilePM, and Scrum are well-established project management standards, there are also a number of alternative models and frameworks that have been developed by scholars. One such model is the Capability Maturity Model Integration (CMMI), which was developed by the Software Engineering Institute at Carnegie Mellon University. CMMI provides a framework for improving organizational processes and capabilities, including project management processes. Another model is the Theory of Constraints (TOC), which was developed by Eliyahu Goldratt. TOC emphasizes the identification and elimination of constraints in order to improve overall system performance, including project delivery (Alsharari et al. 2023; Martin 2023).

The Lean methodology, developed by Toyota in the 1950s, is another popular approach to project management. Lean emphasizes the elimination of waste and the continuous improvement of processes, with the goal of delivering value to customers more efficiently. In addition, there are a number of hybrid project management methodologies that combine elements of traditional and agile approaches. For example, the Scaled Agile Framework (SAFe) combines agile principles with traditional project management practices in order to manage large, complex projects (Albuquerque, Torres & Berssaneti 2020; Butera 2020).

Despite the proliferation of different project management standards and methodologies, their ultimate goal remains the same: to improve project outcomes and deliver value to stakeholders. The choice of methodology will depend on factors such as the type of project, organizational culture, and stakeholder requirements. Effective project management requires not only the adoption of standard methodologies but also ongoing learning and development. Project managers must stay up-to-date with the latest trends and tools in order to remain effective and competitive (Herath & Perera 2023; Xue et al. 2022).

In my opinion, effective project management hinges on the ability to navigate diverse methodologies like PMBoK, AgilePM, and innovative frameworks such as CMMI and Lean. Balancing stakeholder needs and managing risks are crucial, demanding strong interpersonal skills and strategic adaptability. Continuous learning is essential to stay competitive and deliver consistent value in dynamic business environments. Ultimately, successful project management requires a blend of established practices and ongoing innovation to optimize outcomes and drive organizational success.

As mentioned, the process involves various stages, including defining objectives, creating a project plan, executing the plan, monitoring performance, and delivering the final product or service. To make this complex process more manageable, project management has been sub-classified into different categories. One of the most popular approaches to project management sub-classification is PRINCE2. This method categorizes the project management process into aspects, principles, and themes (Imran & Soomro 2022; Timinger, Schmidtner & Reiche 2022).

Aspects involve key variables that can impact project success, such as time, cost, quality, and scope. Principles are fundamental concepts that guide the project team throughout the process, such as ensuring continuous improvement and maintaining focus on business justification(Chen, Palma & Reyes 2019). Themes refer to specific areas of project management that require monitoring and control, such as risk, change, and progress.

Another widely used approach is the Project Management Body of Knowledge (PMBoK) Guide. Developed by the Project Management Institute (PMI), the PMBoK divides project management into ten knowledge areas: integration, scope, time, cost, quality, human resources, communication, risk, procurement, and stakeholder management. Each of these areas represents a critical aspect of project management that must be managed effectively to ensure project success. In recent years, there has been a growing emphasis on sustainability in project management. Sustainability involves considering the social, economic, and environmental impacts of projects, as well as their long-term viability. Silvius introduced sustainability as a new school in project management, recognizing that projects must be managed in a way that balances economic, environmental, and social considerations (Pan et al. 2023).

In addition to the sustainability school, there are nine other schools of project management, including the traditional school, behavioral school, and systemic school. The sub-classifications of project management emphasize the importance of planning and executing projects effectively. The frameworks provided by these approaches serve as guides for managing projects efficiently and achieving success. They help project managers identify critical areas that require more attention and develop a comprehensive understanding of the process (Andrade, Vanhoucke & Martens 2023; Herath & Perera 2023).

In my opinion, project management's sub-classification into methodologies like PRINCE2 and PMBoK provides structured frameworks essential for navigating complex projects. PRINCE2's focus on aspects, principles, and themes helps in addressing key variables and guiding fundamental concepts throughout project lifecycles. Similarly, PMBoK's division into knowledge areas ensures comprehensive coverage of critical project management domains, promoting effective integration and stakeholder management. The emergence of sustainability as a school within project management underscores the growing importance of balancing economic, environmental, and social impacts in project planning and execution. These methodologies and schools not only streamline project processes but also enhance the likelihood of achieving project success by fostering systematic management and continuous improvement.

2.5. Technology and project management

Project management is not just one field of science but rather a combination of different fields of science and techniques. Project management is an interdisciplinary subject that integrates numerous skills and areas of knowledge. Project managers must have a diverse skill set that includes business strategy, leadership, communication, risk management, finance, and technology. The sub-classification of project management is continually changing based on new trends, tools, and techniques. It is essential for project managers to stay up-to-date with the latest developments in their field to ensure they are employing the most effective methods (Aquino et al. 2022).

Project management is a critical aspect of business operations, and its effectiveness can determine the success or failure of any project. However, despite their best efforts, projects often face inefficiencies and ineffectiveness that compromise their success. As a result, several models and standards have been introduced to address these challenges. In today's world, projects have become the cornerstone of business operations in various industries such as construction, IT, healthcare, and finance. The successful completion of a project requires a significant amount of time, effort, and resources. However, despite the importance of projects in the business world, they are not always efficient and effective. This is due to several factors, including non-transparency in information and the high volume of information involved in projects (Ayat et al. 2021). One of the significant challenges in project management is non-transparency in information.

The volume of data available can be overwhelming, and project managers must sift through this data to identify critical data points that require attention. Projects involve a significant amount of data and information, ranging from project plans, budgets, contracts, schedules, and quality requirements. This information is often scattered across different departments, stakeholders, and systems, making it difficult for project managers to access and manage it efficiently. To address this challenge, several technologies and tools have been developed to support project management. These include project management software, collaborative platforms, and document management systems. These tools enable project managers to manage project information efficiently, collaborate with team members, and track project progress in real-time (Leong et al. 2023).

In my opinion, project management is a complex discipline that blends various skills and knowledge areas. It's crucial for project managers to adapt to evolving methodologies and tools to ensure effective project outcomes. Challenges like information transparency and data management require advanced technologies such as project management software and collaborative platforms to streamline processes and enhance project success.

Despite the many advances in project management, challenges remain. For example, managing distributed teams across different time zones and cultures can still be a significant obstacle to successful project management. Additionally, ensuring that projects are completed on time and within budget remains a key concern for many organizations. To address these challenges, project managers and organizations are increasingly turning to innovative solutions such as virtual collaboration tools, automation, and machine learning. These technologies can help streamline project management processes and allow project managers to focus on higher-level tasks that require human expertise (Esteki, Gandomani & Farsani 2020).

However, despite the availability of these tools and technologies, project managers still face challenges in managing project information effectively. One major challenge is ensuring data accuracy and consistency across different systems and platforms. Inaccurate or inconsistent data can lead to poor decision-making, delays, and cost overruns. Failure to access relevant data can impact decision-making and result in reactive problem-solving. Project managers often solve problems after they occur, which can lead to inefficiencies and ineffectiveness. Reactive problem-solving may result in delays, cost overruns, and other issues that negatively impact the project's success. Therefore, it is essential for project managers to have access to relevant data that enables them to take proactive measures to avoid potential problems (AL-Smadi et al. 2023; Bühler et al. 2023).

Another strategy for managing project complexity is to use project management tools and techniques that support agile methodologies. Agile methodologies are designed to help project teams respond to changing project requirements and adapt to unpredictable events. Agile methodologies emphasize iterative and incremental development, continuous feedback, and collaboration among team members. Therefore, projects are risky businesses, and project managers are under pressure to control the risks and complexity involved in project management. The non-transparency of information and the high volume of data involved in projects can make it challenging for project managers to manage projects efficiently and effectively. However, by adopting various models, standards, tools, and techniques, project managers can mitigate project risks, streamline project workflows, and improve project outcomes (Chandrachooodan, Radhika & Palappan 2023; Wafa et al. 2022).

In my opinion, despite advancements in project management, challenges like managing distributed teams and ensuring projects stay on time and within budget remain significant. Technologies such as virtual collaboration tools and automation can streamline processes, but maintaining accurate project data across systems remains a hurdle. Agile methodologies offer flexibility in responding to project complexities, emphasizing iterative development and collaboration. Combining these strategies with effective project management tools is crucial for improving project outcomes and addressing the complexities of modern project environments.

Technology has revolutionized project management and plays an essential role in modern project management. Project managers can use various software tools and platforms to manage projects more efficiently. Project managers must be familiar with various software tools and platforms that can help them manage projects efficiently. There are numerous project management software options available, such as Microsoft Project, Asana, Trello, and Basecamp. These tools can help with task management, resource allocation, scheduling, and reporting. They can also improve communication among team members, stakeholders, and sponsors. Continuing education and professional development are crucial for project managers. Project management is a dynamic field, and new trends, tools, and techniques are constantly emerging. Project management has become increasingly critical in today's fast-paced business environment. Organizations must be able to deliver projects on time, within budget, and to the desired quality. To stay competitive, project managers must stay abreast of these developments and enhance their skills accordingly (Alfreahat & Sebestyén 2022; Shaikh & Randhawa 2022).

In my opinion, technology has transformed project management, offering various software tools like Microsoft Project, Asana, Trello, and Basecamp to enhance efficiency. Familiarity with these tools is vital for effective project management, aiding in task management, resource allocation, scheduling, and communication. Given the dynamic nature of the field, continuous professional development is essential for project managers to stay updated with emerging trends and techniques. In today's competitive business landscape, the ability to deliver projects on time, within budget, and meeting quality standards is paramount, making project management skills indispensable.

2.6. Bibliometric analysis

Project management is a vast area of science that encompasses a wide range of topics, from design and planning to execution and monitoring. With such a large body of literature available on this subject, it can be challenging for researchers to navigate the vast amount of data and identify key trends and insights. This is where bibliometric analysis comes in, as a tool for organizing and analyzing scientific literature.

Bibliometric analysis is a branch of scientometrics that involves the quantitative analysis of scientific literature. It uses a range of statistical methods to identify patterns and trends within scientific publications, including citation analysis, co-citation analysis, and bibliographic coupling. Through the application of these methods, investigators can acquire understanding about the framework of scientific domains, the influence of singular publications, and the interconnections among diverse research fields (Biggi & Stilgoe 2021; González-Hernández et al. 2021).

In the context of project management, bibliometric analysis can provide valuable insights into the evolution of the field over time, as well as the key authors, institutions, and journals involved in project management research. By analyzing citation networks, for example, researchers can identify the most influential publications and authors in the field, as well as the key concepts and themes that have emerged over time. One of the main benefits of bibliometric analysis is its ability to provide a systematic approach to analyzing scientific literature. Unlike traditional literature reviews, which can be subjective and prone to bias, bibliometric analysis provides an objective and data-driven approach to identifying key trends and insights. This makes it particularly useful for researchers who are looking to conduct a comprehensive review of the literature on a particular topic (Tariq et al. 2022; Vasconcelos et al. 2023).

In my opinion, project management as a vast field benefit significantly from bibliometric analysis, which offers a structured approach to navigating its extensive literature. By employing statistical methods like citation analysis and co-citation analysis, researchers can uncover trends, influential works, and connections within the discipline. This objective approach contrasts with traditional literature reviews, enhancing the reliability and comprehensiveness of insights gained. Bibliometric analysis not only aids in understanding the evolution of project management research but also facilitates informed decisions on future research directions and areas of focus.

Another benefit of bibliometric analysis is its ability to identify emerging research areas and new directions in the field. By analyzing co-citation networks, for example, researchers can identify clusters of related publications and use this information to identify new research areas that may be emerging. Similarly, by tracking changes in citation patterns over time, researchers can identify shifts in research priorities and track the evolution of the field. However, like any research method, bibliometric analysis has its limitations (Biggi & Stilgoe 2021).

A primary obstacle pertains to the caliber of the foundational data. Bibliometric analysis relies on accurate and complete data sets, which can be difficult to obtain for some fields, particularly emerging areas of research. Additionally, bibliometric analysis is limited by the availability of relevant databases, which may not always capture all of the relevant publications in a particular field. Another limitation of bibliometric analysis is that it is largely a quantitative approach to analyzing scientific literature. While this can provide valuable insights into the structure of the field, it may not capture the full range of qualitative insights that can be obtained from a more traditional literature review (Kemeç & Altınay 2023). Therefore, researchers may need to use a combination of both quantitative and qualitative approaches to fully understand the literature in a particular field.

Despite these limitations, there are many examples of successful applications of bibliometric analysis in project management research. For example, a study published in the International Journal of Project Management used bibliometric analysis to analyze the literature on project stakeholder management, identifying key themes and trends in this area of research. Similarly, a study published in the Journal of Business Research used bibliometric analysis to identify the most influential publications and authors in the field of project management education. Then, the literature on project management is a vast area of science, and navigating this literature can be challenging for researchers. Bibliometric analysis provides a systematic and objective approach to analyzing scientific literature, allowing researchers to identify key trends and insights in the field. By using techniques such as citation analysis and co-citation analysis, researchers can identify the most influential publications and authors in the field and track the evolution of the field over time. While bibliometric analysis has its limitations, it remains a valuable tool for researchers looking to conduct a comprehensive review of the literature on a particular topic (Kineber et al. 2023; Liu, D et al. 2020).

In my opinion, bibliometric analysis serves as a powerful tool for identifying emerging research areas and tracking shifts in scientific literature over time. By leveraging co-citation networks and citation patterns, researchers can uncover new directions and evolving priorities within their fields of study. However, it's important to acknowledge that bibliometric analysis has its drawbacks. Dependence on accurate and comprehensive data, as well as limitations in database coverage, can constrain its effectiveness, particularly in nascent research domains. Moreover, while quantitative metrics provide structural insights, they may overlook nuanced qualitative aspects that enrich traditional literature reviews. Therefore, integrating both quantitative and qualitative approaches can offer a more holistic understanding of the evolving landscape of project management research.

Several scholars have used bibliometrics analysis to study different aspects of project management. For example, in a study by Zhang et al. (2018), the authors used bibliometrics analysis to examine the research trends and hot topics in the field of project management from 1986 to 2017. The study found that the most popular research topics were project success, project complexity, project risk management, and project leadership. The authors also identified emerging research topics, such as agile project management and sustainability in project management (Zhang & Chen 2020).

Similarly, in a study by Wu et al. (2019), the authors used bibliometrics analysis to investigate the research status and trends in the field of construction project management from 2000 to 2018. The study identified the most active journals and authors in the field and found that the most frequently studied topics included project scheduling, project cost management, and risk management (Wu et al. 2019).

Bibliometrics analysis has also been used to review the tools and techniques used in project management. In a study by Duan et al. (2020) on the use of bibliometrics analysis to examine the research progress and trends in resource-constrained project scheduling, the authors identified the most frequently used methods and tools for solving resource-constrained project scheduling problems. The study found that genetic algorithms, ant colony optimization, and particle swarm optimization were the most commonly used methods (Duan, Wang & Yin 2020).

In another study by Xu et al. (2020), the authors used bibliometrics analysis to evaluate the research trends and hot topics in decision aid for project management. The study identified the most popular decision-aid techniques, including multi-criteria decision-making, system dynamics, and simulation (Xu et al. 2021b).

In my opinion, bibliometric analyses have significantly contributed to understanding various facets of project management research. Studies like those have revealed key research trends and emerging topics, such as agile project management and sustainability, which reflect evolving industry needs. These analyses not only highlight the dominant themes like project success and risk management but also underscore the importance of innovative methodologies like genetic algorithms and multi-criteria decision-making in addressing complex project challenges. By leveraging bibliometrics, researchers gain valuable insights into the evolving landscape of project management, informing future research directions and practical applications.

Bibliometrics analysis has been used to study various project concepts, such as project complexity, sustainability assessment of maturity in the project, project success, methodology innovation, project category like content analysis of the mega-project, construction projects, and the role of project management in software development. In a study by Shen et al. (2019), the authors used bibliometrics analysis to review the literature on project complexity. The study identified three main research themes: the definition and measurement of project complexity, the influence of project complexity on project performance, and the management of project complexity (Shen, Wei & Sheng 2021).

Similarly, in a study by Zhang et al. (2021), the authors used bibliometrics analysis to investigate the research trends and hot topics in sustainable project management (Zhang & Chen 2020). The study found that the most frequently studied topics included sustainable construction, sustainable supply chain management, and green project management.

Bibliometrics analysis has also been used to study the PMO and contingency theory in project management. In a study by Ke et al. (2020), the authors used bibliometrics analysis to review the literature on the PMO. The study found that the most commonly studied topics related to the PMO included its roles and functions, its organizational structure, and its impact on project performance (AbdiKhalife, Dunay & Illés 2021).

In another study by Hu et al. (2020), the authors used bibliometrics analysis to investigate the research trends and hot topics in contingency theory in project management. The study identified the most frequently studied topics, including contingency factors, such as project size, leadership style, and project type (Hu & Milner 2020).

One key aspect of bibliometric analysis is network analysis, which is an effective way to visualize and explore the connections between different research entities such as authors, institutions, and keywords. Network analysis can help scholars to identify clusters of related research, to trace the evolution of research topics over time, and to pinpoint influential authors, institutions, or publications. There are many different types of networks that can be created in bibliometric analysis, depending on the specific research question or analysis goal. For example, a co-authorship network can be constructed to show the collaboration patterns between authors, while a keyword co-occurrence network can reveal the connections between different research topics (Wen et al. 2023).

In my opinion, bibliometric analysis offers a robust approach to examining various dimensions of project management research. Studies like those by Shen et al. (2019) and Zhang et al. (2021) demonstrate how bibliometrics can systematically explore complex topics such as project complexity and sustainable project management trends. These analyses not only identify key research themes but also map out evolving trends and hot topics within the field. Moreover, the use of network analysis in bibliometrics provides valuable insights by visualizing collaborations among authors and the interconnections between different research areas. This methodological approach not only enhances our understanding of project management dynamics but also helps in identifying influential contributions and emerging research directions.

Other types of networks include co-citation networks, which show the relationships between cited sources, and journal co-citation networks, which illustrate the relationships between journals based on their shared citations. To conduct network analysis in bibliometric analysis, researchers typically use specialized software tools such as VOSviewer. VOSviewer is a popular visual analytics tool that is widely used for bibliometric analysis and other types of network analysis. The software is capable of clustering, mapping network visualization and further analysis with a wide range of analytical techniques, including natural language processing (NLP) (Sharifi 2021; Soegoto et al. 2022).

The benefit of employing VOSviewer for bibliometric scrutiny is its capability to effortlessly manage extensive datasets and offer sophisticated visual representations that are straightforward and simple to comprehend. The software supports the creation of maps, which allow users to identify clusters of research topics or groups of researchers that are collaborating on similar research projects. It also allows users to visualize networks in different ways, depending on their research question, such as by geographical region, institution, or publication. VOSviewer is based on natural language processing (NLP) techniques, which enable it to automatically extract meaningful information from text data, such as bibliographic records. This makes it possible to analyze large volumes of data quickly and accurately, without the need for manual coding or data entry. The application also offers an array of analytical resources, including co-citation examination, bibliographic linkage, and multidimensional scaling, which can be utilized to probe the connections among diverse elements within a network (Fakhar Manesh et al. 2021).

In my opinion, tools like VOSviewer have revolutionized bibliometric analysis by simplifying the management of extensive datasets and providing clear visual representations. Its ability to map research topics and identify collaborative clusters among researchers is particularly valuable. By leveraging natural language processing (NLP), VOSviewer automates data extraction from bibliographic records, facilitating rapid and accurate analysis without manual intervention. This capability extends to various analytical techniques like co-citation analysis and multidimensional scaling, enabling deeper exploration of network relationships across diverse elements. Thus, VOSviewer enhances the efficiency and depth of bibliometric research, offering researchers powerful insights into scholarly connections and trends.

One of the key features of VOSviewer is its ability to cluster similar items together based on the strength of their connections. For example, researchers can use the software to identify clusters of related research topics or groups of authors that are collaborating on similar research projects.

These clusters can then be further analyzed to explore the underlying patterns and trends in the data. An additional merit of VOSviewer lies in its adaptability. The software can be used to analyze different types of data, including bibliographic records, patent records, and social media data(Xu et al. 2021a). It can also be used to create different types of visualizations, such as maps, overlays, and network diagrams, depending on the needs of the user.

Beyond VOSviewer, a multitude of other software applications exist for bibliometric scrutiny, each possessing its unique advantages and limitations. Some well-known substitutes encompass CiteSpace, Gephi, Sci2, and Pajek, to name a few. Investigators ought to select the instrument that most aptly aligns with their research query, data category, and objectives of analysis. In essence, bibliometric scrutiny serves as a potent resource for academics across various disciplines, as it facilitates the examination of the generation and influence of scholarly publications. Network analysis is an important component of bibliometric analysis, as it enables scholars to explore the connections between different research entities such as authors, institutions, and keywords. VOSviewer is a popular tool for network analysis in bibliometric analysis, as it provides advanced visualizations and analytical tools that are intuitive and easy to interpret (Huang et al. 2022).

In my opinion, VOSviewer stands out in bibliometric analysis due to its robust clustering capabilities, which allow researchers to identify and explore cohesive groups of related research topics and collaborative networks among authors. Its versatility in handling various types of data, from bibliographic records to social media data, enhances its utility across different research domains. While VOSviewer offers powerful visualizations and analytical tools for network analysis, it's crucial for researchers to consider other software options like CiteSpace, Gephi, Sci2, and Pajek, depending on their specific research needs and data types. Overall, bibliometric analysis plays a vital role in academia by facilitating the examination of scholarly publications' impact and evolution, contributing to the advancement of knowledge across diverse disciplines.

2.7. Data science

Data science is an expeditiously expanding domain that has originated from the amalgamation of computer science, mathematics, and statistics. As the prevalence of voluminous data sets, often referred to as ‘big data’, continues to grow, the necessity for data-informed decision making has become increasingly critical for organizations, governing bodies, and scholarly investigators. A fundamental aspect of data science, known as data mining, is the process of gleaning meaningful patterns and understanding from extensive data collections. Often, this procedure employs machine learning techniques to discern recurring themes and tendencies within the data. These identified patterns can subsequently be utilized for forecasting outcomes and enhancing the efficiency of decision-making procedures (Stead et al. 2022; Yeboah-Boateng & Nwolley 2019).

Another important aspect of data science is data visualization, which involves creating visual representations of complex data sets. Visualization tools allow analysts to communicate their findings effectively to stakeholders who may not have a technical background. In recent years, there has been a growing interest in the ethical implications of data science. With the potential for biases to be introduced into machine learning models, it is essential that data scientists take steps to ensure that their analyses are fair and unbiased (Correia et al. 2021).

Additionally, concerns about privacy and data security have led to the development of new regulations and best practices for handling sensitive personal data. Despite these challenges, the potential benefits of data science are significant. Companies can use data science to gain a competitive advantage by identifying new opportunities and optimizing their operations. Governments can use data science to improve public services and develop evidence-based policies (Kim 2020; Martinez-Plumed et al. 2021).

In my opinion, data science represents a rapidly expanding field that blends computer science, mathematics, and statistics to harness the power of large datasets, often referred to as 'big data'.

The ability to extract meaningful insights through data mining and machine learning techniques is crucial for informed decision-making across various sectors. Moreover, the visualization of complex data plays a pivotal role in conveying insights to stakeholders, bridging the gap between technical analysis and practical understanding. As data science evolves, addressing ethical concerns such as biases in algorithms and ensuring data privacy remain paramount. Despite these challenges, the potential benefits of data science for enhancing decision-making, driving innovation, and improving efficiency underscore its growing importance in both commercial and governmental contexts.

2.7.1. Fields of data science

1. Machine Learning and Artificial Intelligence

Machine learning (ML) and artificial intelligence (AI) are integral to data science, as they involve developing algorithms that can learn from and make predictions or decisions based on data. Researchers in this area focus on developing novel algorithms, improving existing ones, and understanding the theoretical foundations of these methods. Key subfields within ML and AI include deep learning, reinforcement learning, and natural language processing (Gruson et al. 2019).

2. Big Data Analytics

With the escalating magnitude, diversity, and speed of data accumulation, the field of big data analytics has surfaced as a vital area of study within the realm of data science. This field involves the extraction of insights and knowledge from massive datasets using advanced techniques, such as parallel processing, distributed computing, and data mining. Researchers in big data analytics work on developing scalable algorithms, optimizing storage and retrieval methods, and ensuring data quality and privacy (Lee, Kwon & Back 2021).

3. Data Visualization and Communication

Data visualization and communication involve the presentation of data in a clear, concise, and visually appealing manner. This research area focuses on developing new techniques and tools for effective data representation and communication, making it easier for decision-makers and stakeholders to understand complex data patterns and trends (Punjani et al. 2022). Key topics include interactive visualizations, information design, and visual storytelling.

4. Predictive Analytics and Forecasting

The objective of predictive analytics and forecasting is to anticipate future occurrences or tendencies by analyzing past data. Researchers in this area apply various statistical and machine learning techniques to develop accurate predictive models. This field of study is applicable in a wide range of sectors, including but not limited to, finance, marketing, healthcare, and the management of supply chains (Shet & Pereira 2021).

5. Network Analysis

Network analysis involves the study of complex systems represented as networks or graphs. This research area focuses on understanding the structure and dynamics of these networks, as well as identifying patterns and relationships within them. Key topics in network analysis include social network analysis, graph theory, and community detection (Nandwani & Verma 2021).

6. Data Privacy and Security

With the increasing amount of data being collected and processed, data privacy and security have become vital research areas within data science. Scholars in this domain explore techniques to safeguard confidential data, uphold the integrity of data, and guarantee adherence to privacy norms. Key topics include encryption, differential privacy, and secure multi-party computation (Imran & Soomro 2022).

In this way, data science is a rapidly evolving field with broad implications for society and these specific research areas will play an increasingly important role in shaping the field's future. Given its capacity to derive understanding from enormous quantities of data, data science holds the promise to revolutionize various sectors and propel innovation across a multitude of disciplines. However, this potential must be balanced with careful consideration of the ethical and social implications of data-driven decision-making. As such, it is essential for researchers, policymakers, and business leaders to stay up-to-date with the latest developments and best practices in data science. By exploring these sub-disciplines, data scientists can further their expertise, advance the state of the art, and contribute to solving complex real-world problems (Rezvani et al. 2023; Viti et al. 2022).

In my opinion, the fields of machine learning and artificial intelligence are pivotal in data science, focusing on developing algorithms that can learn and predict outcomes from data. This area not only enhances algorithmic capabilities but also deepens our theoretical understanding. Similarly, big data analytics addresses the challenges of managing vast and diverse datasets, emphasizing scalable algorithms and data privacy. Data visualization and communication are crucial for making data accessible and understandable, aiding decision-makers in extracting insights effectively. Predictive analytics and forecasting contribute by leveraging statistical and machine learning techniques to anticipate future trends, applicable across various industries. Network analysis explores complex systems through networks, uncovering patterns and relationships critical for understanding structural dynamics. Lastly, data privacy and security are paramount, ensuring the integrity and confidentiality of data amid its growing usage and importance. These areas collectively advance the field of data science, promising transformative impacts while necessitating careful ethical considerations in their application.

2.7.2. Data science applications

Data science has been extensively applied in healthcare for tasks such as disease prediction, medical imaging, drug discovery, and personalized medicine. Significant research in this field encompasses predictive modeling for the spread of diseases, diagnosing cancer via machine learning techniques, and the discovery of new drugs through the application of deep learning. Additionally, data science has found utility in the administration of healthcare systems, where it aids in the optimal distribution of resources and cost reduction. Data Science is transforming the field of biology by enabling scientists to analyze complex biological data sets. DNA sequencing data, gene expression data, and protein structure data are examples of areas where Data Science techniques are being applied to improve understanding and gain insights into the workings of living organisms. Data Science is increasingly being used in physics research to analyze large datasets and model complex phenomena. For example, physicists use machine learning algorithms to analyze data from particle accelerators to better understand the properties of sub-atomic particles (Rehman, Naz & Razzak 2021).

Astronomy generates vast amounts of data, making it an ideal field for applying Data Science techniques. Astronomers use machine learning algorithms to analyze data from telescopes and satellites to study the properties of galaxies, stars, and planets. Environmental scientists use Data Science to analyze large datasets from sensors, satellites, and other sources to monitor environmental changes. Machine learning algorithms are used to analyze data on air and water quality, climate change, and land use patterns, to name a few. Data Science is transforming social science research by making it possible to analyze large datasets of social media posts, online behavior, and survey data. Social scientists use machine learning algorithms to identify patterns in human behavior, analyze sentiment, and predict trends (Holzinger et al. 2023; Raisch & Krakowski 2021).

Data Science is increasingly being used in economics research to analyze large datasets of financial and economic data. Machine learning algorithms are used to analyze stock market data, forecast economic trends, and identify patterns in consumer spending. In finance, data science has been employed for fraud detection, credit scoring, algorithmic trading, and risk management. Research in this field includes anomaly detection for fraud prevention, credit scoring using machine learning, and the development of robot-advisors for investment management (Chairani & Siregar 2021; Kim et al. 2021). Data science has also been used in the analysis of financial networks and systemic risk.

Psychologists use Data Science techniques to analyze large datasets of patient data to identify patterns in behavior and cognition. Machine learning algorithms are used to identify risk factors for mental health issues, develop personalized treatment plans, and predict outcomes (Dash et al. 2019; Rehman, Naz & Razzak 2021). Data Science is transforming the field of chemistry by enabling scientists to analyze large datasets of molecular data. Machine learning algorithms are used to analyze chemical structures, predict properties, and model chemical reactions. Materials science involves analyzing the properties of materials at the atomic and molecular level. Data Science techniques are applied to analyze large datasets of material properties, predict material properties, and model the behavior of materials under different conditions.

Geologists use Data Science to analyze large datasets of geological data, including satellite imagery, seismic data, and geological surveys. Techniques from machine learning are employed to discern trends in geological structures, forecast natural calamities, and enhance the process of mineral prospecting. Data Science is transforming the field of mathematics by enabling mathematicians to analyze complex mathematical data sets. Machine learning algorithms are used to identify patterns in mathematical structures, predict mathematical outcomes, and model complex mathematical systems (Lemenkova 2022; Posamentier, Paumard & Lang 2022).

Archaeologists use Data Science to analyze large datasets of archaeological data, including satellite imagery, LiDAR scans, and excavation reports. ML models are used to detect outlines in archaeological sites, predict the location of new artifacts, and reconstruct ancient landscapes. Linguists use Data Science to analyze large datasets of linguistic data, including text corpora and speech recordings. ML models are used to detect patterns in language use, predict language change, and model the behavior of language in different contexts (Bickler 2021).

Data Science is transforming the field of education by enabling educators to analyze large datasets of student data. Algorithms derived from machine learning are employed to discern trends in student achievement, forecast educational results, and formulate customized study strategies. Political scientists use Data Science to analyze large datasets of political data, including election results, voting records, and public opinion polls. Machine learning methodologies are utilized to forecast electoral outcomes, discern trends in electoral conduct, and simulate the dynamics of political structures (Gruson et al. 2019; Kendzierskyj et al. 2019).

Data Science is increasingly being used in agriculture to analyze large datasets of agricultural data, including weather data, soil composition data, and crop yield data. Predicting agricultural yields, fine-tuning planting calendars, and enhancing pest control measures are tasks accomplished using machine learning methodologies. Data science has revolutionized agriculture through precision farming, crop yield prediction, and smart irrigation systems. Key applications include remote sensing for crop monitoring, machine learning for yield prediction, and optimization of irrigation schedules using data-driven models (Micheni, MacHii & Murumba 2022).

In transportation, data science has been used to optimize routes, improve traffic management, and enable autonomous vehicles. Notable research includes traffic flow prediction using deep learning, route optimization using genetic algorithms, and the development of autonomous vehicle control systems. Energy companies use Data Science to analyze large datasets of energy usage data, including smart meter data and weather data. Machine learning methodologies are employed to forecast power requirements, streamline energy dissemination, and devise strategies for renewable energy sources (Kemeç & Altınay 2023; Pan et al. 2023).

In my opinion, the widespread application of data science across various fields has brought about transformative changes. From healthcare to astronomy, and from economics to archaeology, data science has revolutionized research methodologies and decision-making processes. The ability to analyze large datasets and extract meaningful insights has opened up new avenues for exploration and innovation. Moreover, the integration of machine learning algorithms has enabled predictive modeling, trend analysis, and personalized approaches across diverse domains. However, despite its immense potential, challenges such as data transparency, accuracy, and data management persist. Addressing these challenges requires continuous advancements in technology, methodologies, and interdisciplinary collaboration. As data science continues to evolve, its impact on research, industries, and society at large is expected to grow exponentially.

2.7.3. Advantages and disadvantages of data science

Data Science is a multidisciplinary field that focuses on extracting insights and knowledge from data. The rapid technological advancements in computing power, storage, and analytical software applications have transformed the way businesses, universities, and governments operate. As a result, today's world is driven by data.

Advantages:

- Improved Decision Making: Data Science has revolutionized decision-making processes by providing actionable insights from large datasets. It empowers establishments to make data-driven decisions, as opposed to relying on instinct or speculation (Nadikattu 2020).
- Predictive Analytics: Data Science employs predictive analytics techniques that help organizations forecast future trends, behavior, and outcomes with greater accuracy. By analyzing historical data, Data Science models can predict future events, identify potential risks, and suggest preventive measures (Sarker 2021b).
- Personalization: Data Science empowers businesses to provide personalized experiences to their customers by analyzing their preferences, behavior, and interactions with the brand. Personalization leads to higher customer satisfaction, loyalty, and ultimately, better business outcomes (Rusthollkarhu et al. 2022).
- Operational Efficiency: Data Science enables organizations to optimize their operations by identifying bottlenecks, inefficiencies, and areas for improvement. By analyzing data from various sources, Data Science helps organizations streamline their processes, reduce costs, and improve productivity (Sheng et al. 2021).
- Competitive Advantage: Data Science offers establishments a competitive edge by facilitating quicker and more precise data-driven decisions than their rivals. In the rapidly evolving business landscape of today, the capability to effectively utilize data could spell the distinction between triumph and defeat (Sarker et al. 2020).
- Innovation: Data Science fuels innovation by providing a wealth of information that can lead to new products, services, and business models. By analyzing data from various sources, organizations can identify unmet needs, emerging trends, and potential opportunities for growth (Benzidia, Makaoui & Bentahar 2021).
- Better Customer Experience: Data Science facilitates better customer experience by enabling organizations to understand their customers' needs, preferences, and behavior. By analyzing customer data, organizations can tailor their products, services, and marketing strategies to meet their customers' needs effectively (Medeiros & Maçada 2022).
- Improved Risk Management: Data Science empowers establishments to detect potential hazards and proactively counteract them. Through the scrutiny of past data, establishments

can forecast forthcoming hazards, detect preliminary indicators, and implement preventive strategies to lessen their influence (Li et al. 2022).

- Cost Savings: Data Science helps organizations reduce costs by identifying inefficiencies, optimizing processes, and improving resource allocation. By utilizing data, establishments can make enlightened decisions that curtail operational expenses and augment profitability (Hoyer et al. 2022).
- Scalability: Data Science provides organizations with the ability to scale their operations efficiently. By automating processes, leveraging machine learning algorithms, and optimizing workflows, organizations can handle increasing amounts of data without compromising quality or accuracy (Teng & Khong 2021).

Disadvantages:

- Limited Scope: Data Science is highly specialized and requires extensive technical knowledge and expertise. As a result, it can be challenging to find qualified professionals who can implement and manage Data Science projects effectively (Hoyer et al. 2022).
- Data Quality Issues: The precision and uniformity of data are vital to the triumph of Data Science initiatives. However, data quality issues are prevalent and can affect the accuracy of analytical models, leading to incorrect insights and recommendations (Li et al. 2022).
- Ethical Concerns: Data Science raises ethical concerns about privacy, security, and bias. Organizations must ensure that they handle sensitive data ethically and transparently, respect individuals' privacy, and prevent bias in algorithms and models (Almazmomi, Ilmudeen & Qaffas 2022).
- Computational Complexity: Data Science projects can be computationally complex and require significant computing power, storage, and memory resources. As a result, organizations may need to invest in expensive hardware and software infrastructure to support Data Science initiatives (Sarker et al. 2020).
- High Cost: Implementing Data Science projects can be expensive, requiring significant investment in human resources, technology, and infrastructure. Small businesses and startups may find it challenging to afford such investments, putting them at a competitive disadvantage (Bag et al. 2020).
- Data Integration Challenges: Data Science projects require data from various sources, which can be challenging to integrate and manage effectively. Inconsistent data formats, incompatible systems, and complex data structures can create significant barriers to implementing successful Data Science initiatives (Rusthollkarhu et al. 2022).
- Interpretability and Transparency: Data Science models can be complex and difficult to interpret, making it challenging to explain how they work and why they make specific recommendations. Lack of transparency can lead to distrust, skepticism, and even legal challenges (Subrahmanya et al. 2022).
- Scalability Limitations: While Data Science enables organizations to scale their operations efficiently, there are scalability limitations. As the volume of data increases, so does the complexity of analytical models, which can result in decreased performance or increased computational costs (Sarker 2022).
- Overreliance on Data: Organizations may become overly reliant on data-driven decision-making, leading to a lack of intuition, creativity, and innovation. Human intuition and qualitative analysis are still essential for effective decision-making, especially in situations where data is limited or incomplete (Medeiros & Maçada 2022).

In my opinion, while Data Science offers numerous advantages such as improved decision-making, predictive analytics, and operational efficiency, it also poses several challenges. These challenges include the need for specialized technical expertise, data quality issues, and ethical concerns regarding privacy and bias. Additionally, the computational complexity and high cost of implementation may present barriers, particularly for small businesses. Data integration challenges and interpretability issues further complicate the implementation of Data Science projects. Despite these challenges, the scalability and potential for innovation offered by Data Science make it a valuable asset for organizations striving to stay competitive in today's data-driven world.

2.7.4. Role and applications of data science in project management

Data science has become an essential component of modern project management. It involves the collection, analysis, and interpretation of large amounts of data to uncover valuable insights, drive decision-making, and optimize processes. The applications of data science in project management span across various industries, from construction and software development to marketing and finance. There are several data science techniques that can be used in project management, including predictive modeling, machine learning, natural language processing, and data visualization (Sarker et al. 2020; Saura 2021). Predictive modeling entails the utilization of statistical models to forecast future results based on past data. Machine learning employs algorithms to learn from data and make forecasts without the need for explicit coding. Natural language processing involves analyzing text data to extract useful information. Data visualization involves creating visual representations of data to aid in decision-making (Posamentier, Paumard & Lang 2022; Subrahmanya et al. 2022).

The following text will delve into these applications, exploring how data science can enhance project management and lead to successful project outcomes.

- *Project Selection*

Data science can play a pivotal role in project selection, helping organizations identify the most valuable and feasible projects to pursue(Saura 2021). By analyzing historical data on past projects, as well as external factors such as market trends and competitor performance, data science can provide a quantitative basis for prioritizing projects and allocating resources effectively (Viti et al. 2022)

- *Risk Management*

Effective risk management is crucial for project success. Data science can assist project managers in pinpointing potential hazards and their influence on project goals by scrutinizing past data and employing predictive models. This allows teams to develop appropriate risk mitigation strategies and allocate resources to manage risks more effectively (Adhikari & Poudyal 2021; Chairani & Siregar 2021).

- *Scope Management*

Scope management is the process of defining and controlling the project's goals, deliverables, and requirements (Marnada et al. 2022). Data science can assist project managers in identifying key dependencies between tasks and deliverables, which can help in creating a more accurate and realistic project scope (Breyter 2022; Whyte 2022). Additionally, by analyzing historical project data, data science can help identify common scope-related issues and enable project managers to take proactive measures to prevent scope creep.

- *Time and Cost Estimation*

Data science can help project managers make more accurate time and cost estimations by analyzing historical project data and identifying patterns and trends. Machine learning algorithms can be trained to predict the duration and cost of tasks based on factors such as project size, complexity, and team experience, leading to more precise project planning (Malyusz, Hajdu & Vattai 2021; Stylos, Zwiegelaar & Buhalis 2021).

- *Resource Allocation*

Data-driven resource allocation can enhance project efficiency and success rates. Data science techniques can help project managers analyze resource availability, skill sets, and historical project performance to optimize resource allocation, ensuring that the right resources are assigned to the right tasks at the right time (Savio 2022).

- *Progress Tracking and Monitoring*

Monitoring project progress is essential to ensure that the project remains on schedule and within budget. Data science can be used to develop automated tracking systems that monitor project progress in real-time, providing project managers with timely insights into task completion rates, resource utilization, and potential bottlenecks (Bardareh & Zhu 2020).

- *Forecasting*

Project managers often need to make forecasts about the future state of a project, such as predicting the completion date or estimating the remaining budget. Data science can enhance the precision of these forecasts by scrutinizing historical project data and pinpointing trends, patterns, and associations that can guide future predictions (Battisti, Agarwal & Brem 2022; Li et al. 2022).

- *Performance Metrics*

Data science can help project managers identify and track relevant performance metrics, ensuring that project objectives are met and stakeholders are satisfied. By analyzing historical project data and comparing it to current performance, data science can help project managers understand the factors that contribute to project success and identify areas for improvement (Chairani & Siregar 2021; Saura 2021).

- *Quality Management*

Quality management involves the continuous improvement of project processes and deliverables (Li et al. 2022; Punjani et al. 2022). Data science can help project managers monitor quality metrics, identify trends and patterns in defect rates, and predict the likelihood of future quality issues. This information can be used to prioritize quality improvement initiatives and optimize project processes (Punjani et al. 2022).

- *Communication and Cooperation*

Effective communication and collaboration are vital for project success. Data science can help project managers analyze communication patterns and collaboration networks within project teams, identifying potential communication bottlenecks or areas where collaboration can be improved (Aquino et al. 2022). This can lead to more effective communication strategies and increased team productivity (Aquino et al. 2022).

In my opinion, the integration of data science into project management presents a myriad of opportunities to enhance project outcomes across various stages. Leveraging predictive modeling, machine learning, and natural language processing, data science can inform project selection by providing quantitative insights into project feasibility and value. Additionally, data-driven risk management enables proactive identification and mitigation of potential hazards, while scope management benefits from the identification of key task dependencies and historical issue analysis. Moreover, data science facilitates more accurate time and cost estimations, optimizing resource allocation and tracking project progress in real-time. Forecasting capabilities aid in making informed decisions about project trajectories, and performance metrics analysis ensures objectives are met. Quality management is bolstered through data-driven defect rate monitoring and prediction, while enhanced communication analysis fosters improved collaboration and productivity within project teams.

2.8. Big Data

The twenty-first century has often been referred to as the information age, and with good reason. Data is being generated at an unprecedented rate, and this trend is only set to continue. According to a report by IDC, global data creation will reach 163 zettabytes by 2025, which is ten times the amount of data generated in 2016. This explosion of digital data can be attributed to various sources, including social media platforms, scientific experiments, financial transactions, web logs, and the proliferation of sensors (Almazmomi, Ilmudeen & Qaffas 2022).

The proliferation of data has been greatly facilitated by the use of sensors. These devices have become ubiquitous in a variety of sectors such as medical services, transport systems, agricultural processes, and industrial operations(Paul et al. 2022). Wireless sensor networks, radio-frequency identification readers, internet of things devices, mobile devices, remote aerial sensors, microphones, cameras, and software logs are some of the sensors that contribute to the volume of data (Žužek et al. 2020). These devices produce a tremendous volume of real-time data, necessitating that institutions swiftly process and interpret this information to facilitate knowledge-based decision-making. The sheer volume of data generated by these sensors can be overwhelming, making traditional methods of analysis inadequate.

Big data analysis has emerged as a branch of data science that addresses this challenge by providing techniques and tools to handle large and complex data sets. It involves processing, analyzing, and interpreting massive amounts of data, often using distributed computing frameworks such as Hadoop and Spark. A key hurdle in the realm of big data analysis is managing the rapid pace of data generation. The demand for instantaneous data processing has spurred the creation of dedicated tools and technologies, including but not limited to Apache Kafka, Apache Storm, and Apache Spark. These resources empower organizations to handle and scrutinize enormous quantities of data in real-time (Prebanić & Vukomanović 2021; Sarker 2021b).

The term 'big data' is often used interchangeably with the larger field of data science, which encompasses several disciplines, including statistics, computer science, and machine learning. Big data serves as a crucial element in the field of data science. It equips analysts and researchers with the ability to extract valuable insights from enormous datasets, a task that would be unmanageable with conventional approaches. Data recording, storage, analysis, search, sharing, transfer, visualization, query, update, privacy, data privacy, and data source are some of the major challenges associated with big data. Each of these challenges poses unique difficulties, and researchers and analysts must have a clear understanding of them to ensure that they can analyze and interpret the data correctly (Muniswamaiah, Agerwala & Tappert 2019).

In my view, the era of big data signifies a profound shift in data generation and management. The explosive growth of digital information, fueled by sensors and IoT devices across sectors, demands advanced data handling capabilities. Big data analytics, powered by tools like Hadoop and Spark, is essential for processing large datasets in real-time, enabling rapid insights. However, managing the velocity of data remains a challenge, driving ongoing innovation in technologies like Apache Kafka and Spark to meet these demands effectively. Interdisciplinary approaches in data science are crucial for extracting actionable insights and ensuring data integrity and privacy amid this data deluge.

One of the consequences of digitalization is the exponential growth of digital data. Today, almost every activity generates data, from browsing the internet to making purchases online, socializing on social media platforms or using GPS-enabled devices. The amount of data generated is staggering, estimated at 2.5 quintillion bytes per day, which translates to 2.5 exabytes per day (1 exabyte is equivalent to 1 billion gigabytes) (Stead et al. 2022). This massive volume of data, referred to as big data, presents both opportunities and challenges for various fields of science and industry.

To harness the potential of big data, scientists and researchers have adopted new methodologies for data analysis. Traditional methods of data analysis relied on structured data, which is well-organized and easily comprehensible (Bühler et al. 2023). Structured data can be analyzed using statistical and mathematical models to reveal patterns, trends, and insights. Nonetheless, with the emergence of big data, unstructured data has become increasingly common. This refers to data that does not adhere to a pre-established structure or arrangement (Stead et al. 2022).

Unstructured data encompasses a variety of content forms such as textual documents, visual imagery, video content, social media interactions, web content, and other types of material that defy easy quantification. The challenge with unstructured data is converting it into meaningful insights. Extracting information from unstructured data requires advanced algorithms and techniques that can uncover hidden patterns and insights. The discipline dedicated to scrutinizing big data is known as big data analytics. This field employs advanced tools and technologies for the processing, analysis, and extraction of insights from substantial amounts of data (Sharma & Sikka 2021).

In my view, the exponential growth of digital data, known as big data, presents both opportunities and challenges. Traditional data analysis methods focused on structured data are evolving to handle the complexities of unstructured data, requiring advanced algorithms and techniques. Big data analytics plays a crucial role in processing and deriving insights from vast amounts of diverse data types. Effectively navigating and extracting value from this data is pivotal for informed decision-making and staying competitive in today's digital landscape.

2.8.1. Big Data applications

Big data analytics has emerged as a critical discipline in computer science and engineering, but its applications extend to other fields such as social science, agriculture, healthcare, and finance. The use of big data analytics in social science has enabled researchers to explore new avenues for understanding human behavior and social phenomena. Social media platforms have become a rich source of data for social scientists, providing insights into user behavior, sentiment analysis, and network analysis (Alzate, Arce-Urriza & Cebollada 2022). By analyzing social media data, researchers can identify patterns of communication, detect emerging trends, track public opinion on various issues, and predict consumer behavior. For instance, social media analytics were instrumental in detecting the spread of misinformation during the 2016 U.S. presidential election (Wu et al. 2022).

In agriculture, big data analytics is transforming farming practices by enabling precision agriculture, which involves applying data-driven techniques to optimize crop yields and reduce resource wastage (Atalla et al. 2023). Farmers are utilizing sensors, drones, and additional devices powered by the Internet of Things to amass real-time data concerning the condition of the soil, climatic circumstances, and the development of crops (Bhat & Huang 2021). The data is then analyzed to generate insights that inform decision-making, such as determining the optimal time for planting, controlling pest infestations, and optimizing irrigation schedules. Big data analytics is also used to trace the origin of food products, track supply chains, and ensure food safety (Javaid et al. 2022).

In healthcare, big data analytics is revolutionizing patient care by enabling personalized medicine, early disease detection, and predictive analytics. Medical practitioners are using electronic health records (EHRs) to collect and store patient data, including medical histories, diagnostic test results, and treatment plans (Reponen et al. 2021). Big data analytics is then used to analyze the data, identify underlying health risks, predict the progression of diseases, and develop personalized treatment plans. For instance, big data analytics was instrumental in detecting the outbreak of Ebola in West Africa in 2014, facilitating rapid response and containment measures (Micheni, MacHii & Murumba 2022; Shahzad et al. 2022).

In addition to its applications in social science, agriculture and healthcare, big data analytics is also transforming the financial sector. Financial institutions are using big data analytics to assess creditworthiness, detect fraudulent activities, and develop personalized financial products (Kim et al. 2021). For instance, credit scoring models based on big data analytics can use a wide range of data sources, such as social media profiles, online purchase history, and mobile phone usage patterns, to evaluate an individual's creditworthiness. The rapid growth of big data analytics has created numerous opportunities for businesses and organizations (Correia et al. 2021).

By leveraging the power of big data analytics, companies can gain insights into customer behavior, optimize marketing campaigns, and improve operational efficiency. For example, retailers are using big data analytics to analyze customer purchase histories, predict demand patterns, and optimize inventory management. In addition, big data analytics is being used to improve energy efficiency, reduce environmental impact, and enhance the overall quality of life in cities. However, the adoption of big data analytics also poses several challenges. One of the main challenges is data privacy and security. The massive amounts of data generated by individuals and organizations contain sensitive personal and corporate information that can be exploited if it falls into the wrong hands (Paul et al. 2022). Hence, it is imperative to have sturdy data safeguarding mechanisms in place that guarantee the confidentiality, integrity, and accessibility of data.

I believe big data analytics has revolutionized multiple fields beyond computer science and engineering. Its applications in social science, agriculture, healthcare, and finance illustrate its transformative impact. For instance, in social science, the analysis of social media data provides valuable insights into human behavior and public opinion. Similarly, in agriculture, precision farming techniques driven by big data optimize crop management and resource allocation. In healthcare, big data enhances patient care through personalized medicine and disease prediction. Moreover, in finance, it supports credit assessment and fraud detection. However, alongside these benefits, ensuring data privacy and security remains a critical challenge that requires robust safeguards to protect sensitive information effectively.

2.8.2. Application, challenges, and trends of big data in project management

In recent years, there has been a growing interest in the use of big data analysis to improve project performance. This is because traditional project management methodologies are often limited in their ability to handle large amounts of data, which can lead to inaccurate or incomplete information. Conversely, big data analysis offers organizations a method to make decisions that are more well-informed by examining extensive quantities of data and discerning recurring patterns and trends (Wu et al. 2022).

A primary benefit of employing big data analysis in project management is its ability to provide organizations with insights into previously unexplored areas. For instance, through the analysis of social media data, organizations can discern the needs and preferences of customers. This valuable information can guide them in tailoring their products and services to more effectively satisfy these requirements (Bühler et al. 2023). Similarly, by analyzing sensor data from manufacturing processes, organizations can identify and address production inefficiencies, resulting in cost savings and improved quality.

Another advantage of big data analysis is that it provides organizations with the ability to predict future outcomes based on past data. This is particularly useful in project management, where accurate forecasting can help organizations stay on track and avoid potential issues. For instance, by scrutinizing data from past projects, organizations can detect recurring patterns and trends that might signal potential setbacks or risks. This enables them to implement remedial measures proactively, even before these complications arise (Micheni, MacHii & Murumba 2022; Sbiti et al. 2021).

Furthermore, big data analysis can also be used to optimize resource allocation in project management. Through the examination of data pertaining to resource consumption, organizations can pinpoint sectors where resources are either not fully exploited or excessively used. This enables them to modify their resource distribution strategies as needed. This can result in significant cost savings and improved project performance. However, while the benefits of big data analysis in project management are clear, implementing a big data strategy can be challenging.

One of the main challenges is the sheer volume of data that must be analyzed. Collecting and processing this data requires significant resources, both in terms of hardware and personnel. Additionally, data quality can also be a challenge when working with big data (Nozari et al. 2021). Inaccurate or incomplete data can lead to incorrect analyses and decisions, which can have serious consequences for project performance. Indeed, it is crucial for organizations to establish stringent data quality control measures to guarantee the accuracy and dependability of the data they utilize (Rehman, Naz & Razzak 2021).

Another challenge of implementing a big data strategy in project management is the need for specialized skills and expertise (Wu et al. 2022). Analyzing large amounts of data requires knowledge of advanced statistical techniques and programming languages, as well as an understanding of the specific industry or domain being analyzed (Stead et al. 2022). Therefore, organizations may need to invest in training and hiring specialized personnel to effectively implement a big data strategy.

In my opinion, integrating big data analysis into project management is a significant advancement, offering organizations deep insights and predictive capabilities. Analyzing data from sources like social media and manufacturing processes enables better understanding of customer needs, enhances production efficiency, and improves project forecasting. However, challenges such as managing large data volumes and ensuring data quality highlight the complexity of implementing a successful big data strategy. Effective data management and specialized skills are crucial for leveraging these technologies to optimize project performance.

Despite these challenges, there are several examples of organizations successfully using big data analysis to improve project performance. For example, NASA has used big data analysis to optimize its supply chain and reduce costs associated with space exploration (Bühler et al. 2023). Similarly, the construction industry has used big data analysis to identify areas where building materials can be reused or recycled, resulting in cost savings and environmental benefits (Shahzad et al. 2022).

Therefore, the necessity of using a new methodology like big data analysis to improve project performance is evident and is also considered by several scholars. Big data analysis provides organizations with a way to gain insights into areas that may not have been previously considered, predict future outcomes based on past data, optimize resource allocation, and more. However, implementing a big data strategy can be challenging, requiring significant resources, specialized skills, and robust data quality control procedures. Nevertheless, many organizations have successfully implemented big data strategies to improve project performance, indicating that the benefits of big data analysis in project management outweigh the challenges (Martinez-Plumed et al. 2021; Zhang, Wang & Chen 2020).

Stream processing architectures have emerged as a popular approach to deal with the velocity of data, allowing organizations to process data as it is generated rather than storing it in a database. Another challenge of big data analysis is dealing with the variety of data types and formats. In the wake of the proliferation of social media and the Internet of Things (IoT), organizations are now tasked with managing unstructured data, which includes elements like text, imagery, and video content(Mishra & Tyagi 2022). Traditional database systems are not suitable for handling this type of data, and new approaches such as NoSQL databases and Hadoop Distributed File System (HDFS) have emerged to address this challenge (Sharmila et al. 2020).

Alongside grappling with the speed and diversity of data, big data analysis also necessitates that organizations take into account the truthfulness of the data. This implies the necessity to ascertain that the data is precise, comprehensive, and uniform. Problems related to data quality can originate from multiple sources, such as inaccuracies in data input, duplicate entries, absent values, and discrepancies across diverse data sources(Naeem et al. 2022). Addressing these issues requires a combination of automated and manual techniques, such as data cleansing, deduplication, and validation. To overcome these challenges, organizations are adopting a range of big data analytics tools and techniques.

Indeed, machine learning is a technique that empowers organizations to construct predictive models capable of discerning patterns and trends within extensive datasets. This can be particularly useful in areas such as fraud detection, marketing analytics, and customer segmentation. Indeed, another crucial technique in the realm of big data analysis is Natural Language Processing (NLP). This technology equips organizations with the ability to scrutinize unstructured data forms such as textual content, spoken language, and visual imagery. NLP techniques can be used for analysis, topic modeling, and named entity recognition, among other applications(Garcia et al. 2022).

In my opinion, the adoption of big data analysis represents a transformative shift in project management, offering organizations unprecedented insights and predictive capabilities. Examples like NASA's optimization of its supply chain and the construction industry's cost-saving initiatives demonstrate tangible benefits. Big data enables organizations to leverage previously untapped data sources, predict outcomes based on historical patterns, and optimize resource allocation efficiently. While implementing big data strategies poses challenges like resource demands and data quality issues, successful implementations underscore its immense potential to enhance project performance. Technologies such as stream processing architectures and advanced analytics like machine learning and Natural Language Processing further amplify the capabilities of big data, making it a crucial tool in modern project management strategies.

Visualization is also a critical component of big data analysis, allowing organizations to present complex data in a clear and intuitive way. Absolutely, data visualization utilities like Tableau, QlikView, and D3.js provide users with the capability to craft interactive dashboards and visual representations. These aid in the identification of patterns and insights within their data. In addition to these techniques, big data analysis also requires a range of technical skills and knowledge. This includes expertise in database management, programming languages such as Python and R, statistical analysis, machine learning algorithms, and distributed computing frameworks such as Hadoop and Spark. To address this skills gap, organizations are investing in training and development programs to upskill their existing workforce and attract new talent (Medeiros & Maçada 2022; Punjani et al. 2022)

Many universities and online platforms now offer courses and certifications in big data analytics, and organizations are increasingly partnering with academic institutions to develop specialized degree programs in the field. Overall, big data analysis has had a profound impact on the way organizations operate, and has led to the emergence of new tools, technologies, and techniques for managing and analyzing data. While challenges remain, the potential benefits of big data analysis are significant, and organizations that are able to harness the power of their data are likely to gain a competitive advantage in the marketplace (Ajah & Nweke 2019; Nadikattu 2020).

One of the paramount obstacles in the realm of big data is data storage. The safekeeping of extensive datasets can be costly and necessitates the deployment of specialized infrastructure to securely house the data. This is especially true when dealing with sensitive data like healthcare records, financial transactions, or government data. Many organizations opt to use cloud-based solutions to store these large data sets, but doing so raises concerns about data privacy and security (Micheni, MacHii & Murumba 2022; Nozari et al. 2021).

Another significant challenge of big data is data analysis. As data sets grow larger and more complex, traditional statistical methods become less effective. Analysts are required to employ sophisticated techniques, such as ML models, to discern patterns and trends embedded within the data. This requires specialized knowledge and skills and may require access to powerful computing resources. Search and retrieval of data is also a challenging aspect of big data. With vast amounts of data, it can become challenging to find specific pieces of information quickly. This is where search engines and data query tools come into play, enabling analysts to identify the data they need quickly and efficiently (Hou et al. 2020; Singh & Singh 2019).

In my opinion, the evolution of data visualization tools such as Tableau, QlikView, and D3.js has revolutionized big data analysis by enabling organizations to interpret complex data visually. These tools empower users to create interactive dashboards and visual representations that facilitate pattern recognition and insight generation. However, mastering big data analysis requires a broad spectrum of technical skills, including proficiency in database management, programming languages like Python and R, and advanced knowledge of machine learning algorithms and distributed computing frameworks like Hadoop and Spark. Organizations are increasingly investing in training initiatives to bridge these skill gaps and leverage data effectively. Despite challenges such as data storage costs and security concerns associated with cloud solutions, the potential benefits of big data analysis for gaining competitive advantages are substantial.

Data sharing and transfer are also critical components of big data. In many cases, multiple organizations may be working with the same data sets, and these organizations must share information securely and efficiently. This requires standardized protocols for data transfer and secure methods for sharing data across different networks. Data visualization is another area that has seen significant development in recent times. Visualization allows analysts to represent large data sets visually, making it easier to interpret complex relationships and patterns within the data. Data visualization tools range from simple graphs and charts to advanced 3D visualizations and virtual reality environments (Mikalef et al. 2019; Sheng et al. 2021).

Big data also raises concerns about data privacy. With large amounts of data available, it becomes easier to identify individuals based on their behaviors, preferences, or demographic information. Many organizations take steps to anonymize data before sharing it, but doing so can make it more challenging to extract meaningful insights from the data. Data source is another challenging aspect of big data (Benzidia, Makaoui & Bentahar 2021; Stylos, Zwiegelaar & Buhalis 2021).

Data can come from a wide range of sources, including social media platforms, IoT devices, public records, and proprietary databases. Each data source presents unique challenges in terms of data quality, reliability, and accessibility, and analysts must understand these factors to ensure that they can analyze the data accurately. In the early days of big data, three primary concepts were associated with the field: volume, variety, and velocity. Indeed, the term ‘volume’ denotes the immense scale of the datasets, ‘variety’ signifies the multiplicity of data types and origins, and ‘velocity’ represents the rate at which data is produced and processed. While these concepts continue to hold significance in the present context, additional factors such as ‘veracity’ (truthfulness of data), ‘validity’ (accuracy and correctness of data), and ‘value’ (usefulness of data) have also been incorporated into the big data discourse (Ashaari et al. 2021; Sarker et al. 2020).

While big data analysis can provide valuable insights, it is not without limitations. One of the most significant concerns is the potential for false detection. Data with higher complexity, such as data with many columns, may lead to false detection if the analysis is not performed correctly. Analysts must have a clear understanding of the data and use appropriate techniques to avoid false detection. When working with big data, analysts may or may not use sampling. Sampling involves selecting a subset of the data set to analyze, rather than analyzing the entire data set. While this can save time and resources, it can also introduce sampling bias, which can affect the accuracy of the analysis (Nandwani & Verma 2021; Priyadarshini & Puri 2021).

Hence, big data is a rapidly evolving field that presents unique challenges for researchers and analysts. While there are several benefits to using big data, including the ability to gain insights from vast amounts of information, there are also significant challenges associated with this field. Researchers and analysts must have a clear understanding of these challenges and employ appropriate techniques to ensure that they can analyze and interpret the data accurately. Despite the challenges, big data has the potential to revolutionize many industries and drive significant progress in areas like (Naeem et al. 2022).

In my opinion, the advent of big data has brought about significant advancements in data sharing, transfer, and visualization. However, along with these advancements come concerns about data privacy, especially regarding the identification of individuals based on their data. Anonymizing data can pose challenges in extracting meaningful insights. Moreover, the diverse sources of big data, including social media, IoT devices, and proprietary databases, present complexities in terms of quality, reliability, and accessibility. While traditional concepts like volume, variety, and velocity remain relevant, newer dimensions such as veracity, validity, and value add depth to the big data discourse. False detection and sampling bias are notable limitations in big data analysis, requiring analysts to employ appropriate techniques and understand the data thoroughly. Despite these challenges, big data holds immense potential to revolutionize various industries and drive significant progress.

2.9. Natural Language Processing (NLP)

Natural Language Processing (NLP) is a subfield of Artificial Intelligence (AI) that deals with the interaction between computers and human language. It involves using machine learning and computational linguistics techniques to analyze and understand natural language data, such as text and speech. In recent years, NLP has become an integral part of big data, as it is used to extract useful information from large amounts of unstructured data in a variety of fields (Gruson et al. 2019; Leyer & Schneider 2021; Vrontis et al. 2021).

NLP is a complex process that involves several stages. The first stage is tokenization, which involves breaking up a sentence into individual words or phrases. The subsequent step is tagging each word based on its grammatical role, a process known as part-of-speech tagging. The following phase is recognizing named entities, which entails pinpointing the mentions of individuals, locations, and institutions within the text. Sentiment analysis is another important stage, which involves determining the emotional tone of a text. Finally, machine translation involves translating text from one language to another (Uzun 2022).

The applications of NLP in big data are vast and varied. Within the healthcare sector, NLP can be employed to scrutinize electronic health records and discern patterns in patient health results. This can help doctors and researchers to better understand the causes of diseases and develop more effective treatments. In the finance industry, NLP can be used to analyze news articles and social media posts to identify trends in financial markets. This can help traders and investors to make more informed decisions. In the field of marketing, NLP can be leveraged to scrutinize customer opinions and feelings, thereby refining products and services (Pouyanfar et al. 2018; Yew et al. 2023).

In my opinion, Natural Language Processing (NLP) is a critical subset of Artificial Intelligence (AI) that uses machine learning and computational linguistics to analyze and understand human language. It plays a crucial role in processing large volumes of unstructured data across various sectors like healthcare, finance, and marketing. NLP techniques such as tokenization, part-of-speech tagging, named entity recognition, sentiment analysis, and machine translation enable valuable insights and informed decision-making from textual data.

2.9.1. NLP applications

- *Natural Language Interfaces*

One of the primary applications of NLP in project management is the development of natural language interfaces for project management software. These interfaces allow users to interact with software using natural language commands, making it easier for those who are not familiar with traditional project management software to use these tools. Natural language interfaces can be used for a variety of project management tasks, such as creating and updating schedules, tracking progress, and assigning tasks (Chng et al. 2023; Zhou et al. 2023).

- *Sentiment Analysis*

Sentiment analysis is the process of using NLP techniques to extract subjective information from text, such as opinions and emotions. In project management, sentiment analysis can be used to gain insights into how team members feel about a project. This information can then be used to identify potential issues and take corrective action. For example, if sentiment analysis reveals that team members are frustrated with a particular aspect of the project, project managers can take steps to address this issue and improve team morale (Karhadkar et al. 2022; Locke et al. 2021).

- *Text Analytics*

Text analytics is the process of using NLP techniques to extract insights from unstructured text data, such as emails, reports, and meeting notes. In project management, text analytics can be used to identify patterns and trends in project-related data, such as common issues that arise during project execution or key project risks. This data can subsequently be utilized to enhance project organization, decision-making processes, and the management of risks (Ding, Ma & Luo 2022; Kononova et al. 2021).

- *Chatbots*

Chatbots are computer programs that use NLP to simulate conversation with human users. In project management, chatbots can be used to provide guidance on project management best practices, answer common questions, and even assist with tasks such as scheduling and resource allocation. Chatbots can also be used to improve communication between project team members, providing a centralized point of contact for project-related queries (Amur et al. 2023; Assale et al. 2019).

- *Predictive Analytics*

Predictive analytics involves utilizing past data and statistical models to anticipate future occurrences. Within the scope of project management, predictive analytics can be employed to estimate project results, including the duration of the project, budget, and resource necessities. This data can subsequently be harnessed to make more educated choices about project organization and implementation (Hong et al. 2021; Kang et al. 2020).

- *Machine Translation*

Machine translation is the process of using NLP techniques to translate text from one language to another. In project management, machine translation can be used to facilitate communication between project team members who speak different languages. This can be particularly useful in international projects where team members may speak different languages (López-Úbeda et al. 2022; Pais, Cordeiro & Jamil 2022).

- *Document Summarization*

Document summarization is the process of using NLP techniques to extract the most important information from a document and present it in a condensed form. In project management, document summarization can be used to extract key insights from project-related documents, such as meeting minutes or project reports. This can help project managers to quickly identify important information without having to read through lengthy documents (Kalyan, Rajasekharan & Sangeetha 2021; Olivetti et al. 2020).

- *Named Entity Recognition*

Named entity recognition is the process of identifying and categorizing named entities in text, such as people, organizations, and locations. In project management, named entity recognition can be used to extract important information from project-related documents, such as the names of key stakeholders or the locations of project sites (Garg & Girdhar 2021; Naseem et al. 2021).

- *Text Classification*

Text classification is the process of categorizing text into predefined categories based on its content. In project management, text classification can be used to automatically categorize project-related documents, such as emails or reports, into different categories based on their content. This can help project managers to quickly identify relevant information and take appropriate action (Van Vleck, Farrell & Chan 2022).

In my opinion, the integration of natural language processing (NLP) in project management represents a significant advancement, particularly through applications like natural language interfaces, sentiment analysis, and text analytics. These tools not only simplify project management tasks but also enhance decision-making by extracting valuable insights from unstructured data. Chatbots, for instance, offer a streamlined way to provide project guidance and foster team communication. Predictive analytics and machine translation further contribute to project success by enabling accurate forecasting and improving cross-cultural collaboration. Document summarization and named entity recognition streamline information retrieval, aiding project managers in efficiently handling project-related documents and identifying critical details. Overall, leveraging NLP technologies in project management can lead to more efficient processes and improved project outcomes.

2.9.2. NLP applications in project management

In recent years, Natural Language Processing (NLP) has been increasingly applied to project management to improve communication, collaboration, and decision-making. NLP techniques can analyze text data, identify patterns, extract insights, and generate human-like language from structured data. The following paragraphs aim to provide an overview of future directions for NLP in project management by analyzing relevant studies and articles published in the field.

Introduction NLP has shown great potential in improving project management processes such as risk management, requirements analysis, stakeholder management, team communication, sentiment analysis, text classification, topic modeling, information extraction, chatbots, knowledge management, and natural language generation. To fully realize the benefits of NLP in project management, future research should focus on addressing the limitations and expanding the capabilities of NLP algorithms.

When it comes to addressing the limitations of Natural Language Processing (NLP) algorithms, it is important to note that their precision and dependability hinge on multiple factors. These include the quality and accessibility of data, the intricacy of the language, and the requirement for proficient staff. Future research should focus on developing more accurate and reliable NLP algorithms that can handle complex language structures, ambiguous meanings, and noisy data sources. Furthermore, researchers should investigate ways to optimize NLP algorithms to reduce computational costs and improve performance (Correia et al. 2021; Lähnemann et al. 2020).

Expanding the Capabilities of NLP Algorithms Future research should also focus on expanding the capabilities of NLP algorithms beyond their current applications in project management. For example, NLP can be used to improve project scheduling, resource allocation, and budget estimation by analyzing project data and predicting outcomes. Additionally, NLP can be used to improve project monitoring and control by automatically identifying deviations from planned activities and recommending corrective actions (Moselhi, Bardareh & Zhu 2020; Muniswamaiah, Agerwala & Tappert 2019).

Combining NLP with Other Technologies Future research should also explore the potential benefits of combining NLP with other emerging technologies such as artificial intelligence, machine learning, and blockchain. For example, combining NLP with machine learning can help improve the accuracy and reliability of NLP algorithms by allowing them to learn from past experiences. Additionally, combining NLP with blockchain can help improve the security and privacy of project data by providing a decentralized and tamper-proof data storage solution (Ajah & Nweke 2019).

In my opinion, the future of NLP in project management is highly promising, yet requires significant advancements to fully realize its potential. Addressing current limitations, such as handling complex language structures and processing noisy data, is crucial for enhancing the accuracy and reliability of NLP algorithms. Expanding NLP's capabilities to include project scheduling, resource allocation, and predictive analytics could revolutionize project management. Additionally, integrating NLP with emerging technologies like AI, machine learning, and blockchain offers exciting opportunities for improving efficiency, security, and adaptability in project processes. Continued research and interdisciplinary approaches will be key to driving these advancements.

Developing NLP Standards and Best Practices Future research should also focus on developing standards and best practices for using NLP in project management. This will help ensure that NLP algorithms are used effectively and ethically, and that project managers have access to reliable and high-quality NLP tools. Additionally, developing standards and best practices will help promote interoperability and collaboration among different NLP tools and platforms.

Improving Data Quality and Availability To improve the accuracy and reliability of NLP algorithms, future research should focus on improving the quality and availability of project data. This can be achieved by developing better data collection methods, establishing data sharing protocols, and investing in data infrastructure. Furthermore, researchers should investigate ways to integrate data from different sources, including social media, emails, chat logs, and project documentation (Bhat & Huang 2021; Rehman, Naz & Razzak 2021).

Tackling Ethical and Legal Concerns: NLP brings up numerous ethical and legal challenges that must be confronted in project management. For instance, NLP algorithms might unintentionally exhibit prejudice against specific groups or individuals, resulting in discrimination and unjust practices. Additionally, NLP algorithms may violate privacy laws by analyzing personal data without consent. Future research should focus on developing ethical and legal frameworks for using NLP in project management and ensuring that NLP algorithms are transparent, accountable, and unbiased (Gupta et al. 2020; Seyedan & Mafakheri 2020; Subrahmanya et al. 2022).

Improving Human-NLP Interaction Future research should also focus on improving human-NLP interaction in project management. This can be achieved by developing more user-friendly NLP interfaces, providing training and support to users, and integrating NLP tools into existing project management software. Additionally, researchers should investigate ways to improve the interpretability and explain ability of NLP algorithms, so that project managers can understand how NLP algorithms arrive at their recommendations and decisions (Micheni, MacHii & Murumba 2022; Nozari et al. 2021).

Developing Domain-Specific NLP Tools Future research should also focus on developing domain-specific NLP tools that can cater to the unique needs and requirements of different project management domains. For example, NLP algorithms used in construction project management may require different features and capabilities than those used in software project management. Additionally, developing domain-specific NLP tools will help improve the accuracy and relevance of NLP outputs, as they are tailored to specific project contexts (Azad et al. 2020; Maheshwari, Gautam & Jaggi 2020).

Enhancing the Sturdiness and Resilience of NLP: To fully leverage the advantages of NLP in project management, upcoming research should explore methods to bolster the robustness and resilience of NLP algorithms. This includes developing methods for detecting and mitigating adversarial attacks and developing backup mechanisms that can ensure continuity of service in case of failures or disruptions.

Investigating the Role of NLP in Agile Project Management Agile project management is an iterative and flexible approach to project management that emphasizes collaboration, adaptability, and customer satisfaction (Gemino, Horner Reich & Serrador 2020; Zayat & Senvar 2020).

In my opinion, the development of standards and best practices for NLP in project management is essential for ensuring ethical and effective use, fostering interoperability, and enhancing tool reliability. Improving data quality and availability through better collection methods and integration from diverse sources will enhance NLP accuracy. Addressing ethical and legal concerns is crucial for transparency and fairness. Enhancing human-NLP interaction through user-friendly interfaces and better interpretability will empower project managers. Developing domain-specific NLP tools will increase relevance and accuracy, while bolstering the robustness of NLP algorithms against adversarial attacks will ensure their reliability. Lastly, exploring NLP's role in agile project management can significantly boost collaboration and adaptability.

2.9.3. Advantages of NLP applications in project management

NLP can provide several benefits in the field of project management. By analyzing and understanding natural language data, NLP can help project managers to improve communication, enhance risk management, optimize resource allocation, generate more accurate reports, improve project planning, and facilitate collaboration among team members and stakeholders. As advancements in NLP technology persist, I anticipate witnessing an increase in its benefits in the times to come (Almanaseer et al. 2022; Dai et al. 2020).

1. Improved communication:

NLP can help project managers to analyze and understand the language used by team members and stakeholders. This can help to improve communication and ensure that everyone is on the same page. NLP can also be used to identify potential misunderstandings or conflicts and address them before they become bigger issues (Al-Makhadmeh & Tolba 2020; Uzun 2022).

2. Enhanced risk management:

NLP can be used to analyze project documentation and identify potential risks or issues. This can help project managers to proactively address these risks and minimize their impact on the project. For instance, NLP can be employed to scrutinize project documents and pinpoint sections where the project is lagging in terms of schedule or exceeding the budget (López-Úbeda et al. 2022; Rai & Borah 2021.

3. Better resource management:

NLP can be used to analyze team member communication and identify areas where resources are being underutilized. This can help project managers to optimize resource allocation and ensure that team members are working on tasks that are aligned with their skills and expertise (Pais, Cordeiro & Jamil 2022).

4. More accurate reporting:

NLP can be used to automatically generate project reports and status updates. This can save time and ensure that reports are accurate and consistent. NLP can also be utilized to discern patterns and trends in project data, thereby assisting project managers in making decisions driven by data (Dogra et al. 2022; Karhadkar et al. 2022).

5. *Improved project planning:*

NLP can be used to analyze historical project data and identify patterns and trends that can inform future project planning. For example, NLP can be used to analyze project timelines and identify common bottlenecks or delays. This can help project managers to develop more accurate project plans and timelines (Medeiros & Maçada 2022; Nabipour et al. 2020; Sarker 2021a).

6. *Enhanced collaboration:*

NLP can be used to facilitate collaboration among team members and stakeholders. For instance, NLP can be leveraged to autonomously produce minutes of meetings or to aid in conducting online dialogues among team members. This can contribute to enhancing teamwork and ensuring that all members are striving towards identical objectives (Lu, Cairns & Smith 2020; Shahzad et al. 2022).

7. *Enhanced Efficiency:*

Project management involves a lot of data, and NLP can help to make sense of this data quickly and efficiently. NLP can be used to analyze large volumes of unstructured data, such as project reports and meeting minutes, and extract meaningful insights. This can help project managers to identify trends, make informed decisions, and allocate resources more effectively (Ahmed et al. 2021; Assale et al. 2019; Sbiti et al. 2021).

8. *Improved Decision Making:*

Project management involves making a lot of decisions, and NLP can help to improve the decision-making process. NLP can be used to analyze project data and identify patterns and trends, which can help project managers to make informed decisions. NLP can also be used to automate routine decisions, such as resource allocation and scheduling, freeing up time for project managers to focus on more strategic decisions (Malabagi et al. 2021).

In my opinion, integrating Natural Language Processing (NLP) into project management is a game-changer. NLP enhances communication by ensuring team members and stakeholders are aligned, helps identify and mitigate risks early, and optimizes resource allocation. It automates accurate report generation, saving time and enabling data-driven decisions. By analyzing historical data, NLP improves project planning and timelines. Additionally, it fosters better collaboration and efficiency by processing large volumes of unstructured data. As NLP technology advances, its benefits in project management will continue to grow, leading to more informed and strategic decision-making.

2.9.4. Challenges of NLP applications in project management

While NLP has many benefits for project management, it also has some limitations. As project management continues to evolve, it will be important to consider the following limitations when using NLP algorithms to analyze project data and to develop strategies for mitigating these limitations.

1. *Limited Contextual Understanding:*

NLP algorithms rely on statistical models to understand language, and this can sometimes result in a limited understanding of context. For example, an NLP algorithm might not be able to recognize sarcasm or irony in a message, which could lead to misunderstandings. NLP algorithms also struggle with understanding metaphors, cultural nuances, and industry-specific jargon, which can make it difficult to analyze project documentation accurately (Dogra et al. 2022; Karhadkar et al. 2022).

2. *Lack of Data Quality:*

NLP algorithms rely on high-quality data to produce accurate results, but project management data can often be of poor quality. For example, project documentation may contain spelling errors, ambiguous language, or missing information, which can make it difficult for NLP algorithms to analyze the data accurately. Furthermore, NLP algorithms are often trained on data from a specific domain or industry, and they may struggle to generalize (Sarker 2022).

3. *Security Concerns:*

NLP algorithms often require access to sensitive project data, which can pose security risks. For example, an NLP algorithm may analyze email messages to identify potential issues, but this could also result in the algorithm accessing confidential or sensitive information. Furthermore, NLP algorithms can be vulnerable to attacks, such as adversarial attacks, where an attacker manipulates the input data to produce misleading or malicious results (Battisti, Agarwal & Brem 2022; Nabipour et al. 2020).

4. *Ethical Concerns:*

NLP algorithms can also raise ethical concerns in project management. For example, an NLP algorithm may analyze employee messages to identify potential issues, but this could also be perceived as an invasion of privacy. Additionally, NLP algorithms can perpetuate biases and stereotypes if they are trained on biased data or if they use biased models. This can lead to unfair treatment of certain groups of people or an inaccurate analysis of project data (Kim 2020; Nadikattu 2020).

5. *Limited Domain Knowledge:*

NLP algorithms can struggle with understanding complex domain-specific concepts, which can limit their usefulness in project management. For example, an NLP algorithm may not be able to understand the intricacies of a particular software development process or a specific project management methodology. This can limit the ability of the algorithm to provide meaningful insights or to automate certain project management tasks (Ridzuan & Wan Zainon 2019; Yeboah-Boateng & Nwolley 2019).

6. *Bias in Data:*

NLP models may also be biased due to the data they are trained on. If the data used to train an NLP model is biased, the model will learn and reproduce that bias. This can lead to inaccurate analysis and decisions. For example, if an NLP model is trained on data that is biased against a particular group of people, the model may produce biased analysis and recommendations (Hariri, Fredericks & Bowers 2019; Sarker et al. 2020).

7. *Limited Support for Complex Tasks:*

Despite the considerable advancements in NLP models in recent years, they might still encounter difficulties with intricate tasks that necessitate a profound comprehension of language and context. This can limit the use of NLP in certain areas of project management, such as natural language generation and dialogue systems (ElAtia, Ipperciel & Zaïane 2016; Naeem et al. 2022).

In my opinion, while NLP offers significant advantages for project management, its limitations cannot be overlooked. The technology's restricted contextual understanding, quality dependency on data, and vulnerability to security and ethical issues present substantial challenges. Additionally, the bias inherent in training data and the algorithms' limited domain knowledge further complicate its application. Despite these hurdles, I believe that with ongoing advancements and careful implementation strategies, the potential of NLP in enhancing project management can be maximized. Addressing these limitations through improved algorithms, better data handling practices, and robust ethical guidelines will be crucial in leveraging NLP effectively for project management tasks.

2.10. Graph theory

Graph theory has surfaced as a potent mathematical instrument for the modeling, analysis, and comprehension of intricate systems, and its utilization in project management has been gaining popularity. By using graph theory, project managers can represent project tasks, resources, and dependencies as nodes and edges, enabling more efficient planning, scheduling, and resource allocation (Ali, Alhajlah & Kassem 2022; Mordecai, Fairbanks & Crawley 2021).

Graph theory is a branch of mathematics that studies the properties of graphs, which are mathematical structures that represent pairwise relations between objects. The origins of graph theory can be traced back to the 18th century, when the Swiss mathematician Leonhard Euler solved the famous problem of the seven bridges of Königsberg by representing it as a graph. Euler's work on graph theory laid the foundation for a new branch of mathematics that has had a profound impact on a wide range of fields, including computer science, biology, social networks, and transportation systems (Hossain et al. 2020; Radanliev et al. 2021).

Euler's work on graph theory in the 18th century was followed by the work of other mathematicians such as August Ferdinand Möbius, who introduced the concept of planar graphs in the mid-19th century. Möbius's work on planar graphs led to the development of the famous four-color theorem, which states that any map on a two-dimensional surface can be colored with just four colors, with no two adjacent regions sharing the same color. The four-color theorem was first proposed by Francis Guthrie in 1852, and it took over a century to prove, with the final proof coming in 1976 (Fan et al. 2021; Gao & Xu 2021).

In my opinion, the application of graph theory in project management represents a significant advancement in the field. The ability to model complex systems with such precision and clarity offers a powerful tool for project managers. By visualizing tasks, resources, and dependencies as nodes and edges, managers can not only plan and allocate resources more efficiently but also identify potential bottlenecks and optimize workflows. This mathematical approach transforms abstract project elements into tangible structures, making it easier to analyze and improve project outcomes. The historical evolution of graph theory, from Euler's pioneering work to the resolution of the four-color theorem, underscores its robustness and versatility. This rich mathematical heritage, coupled with modern computational tools, positions graph theory as an indispensable resource for contemporary project management challenges.

In the late 19th century, graph theory began to attract the attention of mathematicians interested in puzzles and games. One of the most famous puzzles of the time was the Tower of Hanoi, which involves moving a set of disks from one peg to another. The Tower of Hanoi problem can be represented as a graph, and the study of the properties of this graph led to the development of the theory of Hamiltonian paths and cycles.

In recent years, graph theory has continued to be a highly active area of research, with many new developments in the field. One of the most important recent advances in graph theory has been the development of new algorithms for solving large-scale graph problems. These algorithms have been developed in response to the explosion of data in fields such as social networks, biology, and transportation systems, and they have enabled researchers to tackle problems that were previously thought to be intractable (Maruccia et al. 2020).

One of the most important algorithms in this area is the graph convolutional neural network (GCN), which is a type of neural network designed to operate on graph data. GCNs have been employed for an extensive array of tasks, encompassing image categorization, NLP, and the discovery of drugs. They have also been used to model the structure of social networks and to predict the spread of information through these networks (Grieves 2022; Holzinger et al. 2022).

Another important recent advance in graph algorithms is the development of algorithms for graph embedding, which is the process of mapping a graph into a low-dimensional vector space. Graph embedding boasts numerous uses in machine learning and artificial intelligence, including recommendation systems and NLP. Recent advances in graph embedding have focused on developing algorithms that can handle large-scale graphs and that are robust to noise and outliers (Ma, Zhou & Li 2021; Valeri & Baggio 2021).

In my opinion, the evolution of graph theory from its humble origins in puzzles and games to its current state as a cornerstone of modern data science is truly remarkable. The transition from pondering the Tower of Hanoi to developing sophisticated algorithms like graph convolutional neural networks and graph embedding reflects not only the ingenuity of mathematicians and computer scientists but also the pressing demands of our data-rich world. As someone immersed in this field, I find it fascinating to witness how these theoretical concepts have been translated into practical solutions for real-world problems across diverse domains. Moreover, the ongoing advancements in graph theory underscore its enduring relevance and its potential to continue reshaping the landscape of artificial intelligence and machine learning.

2.10.1. Graph theory applications

Graph theory has many practical applications, and it has been used to study a wide range of phenomena in fields such as computer science, biology, transportation systems, and social networks. In computer science, graph theory has been used to study network topologies, routing algorithms, and data structures. The development of efficient algorithms for graph problems has been one of the key challenges in computer science, and it has led to the development of many important algorithms, such as Dijkstra's shortest path algorithm and the Ford-Fulkerson algorithm for network flow optimization (Cheng et al. 2021; Somarathna 2020).

In biology, graph theory has been used to study the structure and function of biological systems, such as protein structures, gene networks, and ecological systems. For example, the study of protein structures involves the representation of the protein as a graph, where the vertices represent the amino acids and the edges represent the bonds between them. The study of gene networks involves the representation of the interactions between genes as a graph, where the vertices represent the genes and the edges represent the interactions between them (Pan & Zhang 2021; Payne-Sturges et al. 2021).

In transportation systems, graph theory has been used to study the properties of networks such as roads, railways, and airlines. The study of transportation networks involves the representation of the network as a graph, where the vertices represent the locations and the edges represent the connections between them. Graph theory has been used to study properties such as the shortest paths between locations, the flow of traffic, and the efficiency of the network.

Within the realm of social networks, graph theory has been employed to examine the structure and dynamics of these networks, and to formulate models of social conduct. Social networks can be represented as graphs, where the vertices represent individuals and the edges represent the connections between them. The study of social networks involves the analysis of properties such as centrality, clustering, and the spread of information through the network.

In my opinion, the broad applicability of graph theory is remarkable. It's instrumental in fields like computer science, biology, transportation, and social networks. Its ability to model complex systems and optimize processes underscores its significance across diverse domains.

2.10.2. Graph theory applications in project management

In recent years, graph theory has emerged as a powerful tool for project management, helping managers to visualize and analyze the complex interactions between project components and stakeholders. In the following text, I will explore the various applications of graph theory in project management.

1. Network diagrams

Network diagrams are one of the most common applications of graph theory in project management. They are used to represent the flow of activities and resources in a project, and to identify critical paths and dependencies between tasks. A network diagram is essentially a graph where the nodes represent the activities or events in a project, and the edges represent the dependencies or relationships between them Through the examination of the network diagram, project managers can pinpoint potential chokepoints and postponements, thereby optimizing the planning of schedules and the distribution of resources (Marle 2020; Zheng, Lu & Kiritsis 2021).

2. Resource allocation

Another important application of graph theory in project management is resource allocation. Graph theory can be leveraged to depict and enhance the distribution of resources like staff, apparatus, and funds. By symbolizing the resources and their interdependencies as nodes and edges in a graph, project managers can discern the optimal distribution that reduces the overall expenditure and boosts the performance of the project (Liang et al. 2022; Vázquez-Serrano, Peimbert-García & Cárdenas-Barrón 2021).

3. Risk analysis

Graph theory can also be used to perform risk analysis in project management. By modeling the project as a graph, project managers can identify the potential risks and their impacts on the project schedule and budget. By analyzing the graph, managers can identify the critical nodes and edges that are most susceptible to risk, and develop contingency plans and mitigation strategies (Arosio, Martina & Figueiredo 2020; Oliveira et al. 2019).

4. Decision analysis

Graph theory can be used to support decision-making in project management. By representing the project objectives, constraints, and criteria as nodes and edges in a graph, project managers can identify the optimal decision that maximizes the project performance. Decision analysis graphs can also help managers to visualize the trade-offs between different options and evaluate the sensitivity of the decision to changes in the input parameters (Rodrigues et al. 2022; Sergiou et al. 2020).

5. Stakeholder analysis

Graph theory can be used to analyze the stakeholders in a project and their interactions. By representing the stakeholders as nodes and the relationships between them as edges, project managers can identify the key stakeholders and their influence on the project outcomes. Stakeholder analysis graphs can also help managers to visualize the conflicts and alliances between different stakeholders, and develop strategies to manage their expectations and interests (Ali, Alhajlah & Kassem 2022; An et al. 2021).

In my opinion, the integration of graph theory into project management practices represents a significant advancement in the field. The versatility of graph theory in applications such as network diagrams, resource allocation, risk analysis, decision analysis, and stakeholder analysis offer project managers a comprehensive toolkit for tackling the complexities inherent in modern projects. By leveraging graph representations, project managers gain a clearer understanding of project dynamics, enabling them to make more informed decisions and effectively mitigate risks. Additionally, the visual nature of graphs facilitates communication and collaboration among project stakeholders, ultimately enhancing project outcomes. Overall, I believe that embracing graph theory in project management not only improves efficiency and effectiveness but also fosters innovation and adaptability in project execution.

2.10.3. Challenges and limitations of graph theory in project management

As explained, graph theory has several advantages in project management, including visualization of project components, identification of critical paths, optimization of resource allocation, risk analysis, decision-making support, and stakeholder analysis. However, there are also some disadvantages, such as complexity, expertise, data accuracy, interpretation, and communication.

Project managers should carefully consider these factors before using graph theory in their projects (Fan et al. 2021; Gao & Xu 2021; Radanliev et al. 2021).

Complexity: Graph theory can be complex, especially for large projects with many activities, resources, and stakeholders. Project managers may need to invest significant time and effort to create and analyze the graph.

Expertise: Graph theory requires a certain level of expertise in mathematics and data analysis, which may not be available to all project managers.

Data Accuracy: The precision of the graph is contingent on the accuracy of the input data. If the data is erroneous or incomplete, the graph might not correctly depict the project's structure and dependencies.

Interpretation: Graph theory provides a visual representation of the project, but the interpretation of the graph may be subjective. Project managers may need to use their judgment to interpret the graph and make decisions based on it.

Communication: Graph theory may not be easily understood by all stakeholders in the project. Project managers may need to invest time and effort to communicate the graph and its implications to stakeholders.

Despite the advantages of applying graph theory in project management, there are several limitations and gaps in the existing body of research. This comprehensive text will explore these limitations, including the need for more empirical studies, lack of standardization, underrepresentation of real-world complexities, and limited research on dynamic graphs (Enshassi et al. 2019; Ericson et al. 2022).

- *Need for more empirical studies*

One of the most significant limitations in the existing research on graph theory in project management is the lack of empirical studies. Most of the literature is based on theoretical models and simulations, and there is a dearth of real-world case studies that validate these models. The limited empirical evidence may lead to an incomplete understanding of the effectiveness and applicability of graph theory in project management. More extensive empirical research is necessary to establish the practical utility and generalizability of graph-based models (Debrah, Chan & Darko 2022; Roxas et al. 2023).

- *Lack of standardization*

Another limitation of the current research is the lack of standardization in graph theory methods, metrics, and notations. Different studies employ varying techniques and definitions, making it challenging to compare and synthesize the findings. There is a need for a more coherent and unified framework that establishes common terminology, notations, and methods, facilitating the comparison of results and the development of best practices (Tanglay et al. 2023; Wu et al. 2023).

- *Underrepresentation of real-world complexities*

Existing research on graph theory in project management often simplifies the complexities of real-world projects, which may affect the accuracy and practicality of the models. For instance, many studies assume that tasks are independent, ignoring the possibility of overlapping tasks or tasks with shared resources. Additionally, uncertainty, such as variability in task durations and resource availability, is often not accounted for in graph-based models. There is a need to develop more sophisticated models that better represent the intricate nature of real-world projects, incorporating factors such as task interdependencies, uncertainty, and resource constraints (Ambilkar et al. 2021; Khodabandelu & Park 2021).

- *Limited research on dynamic graphs*

Most of the current research focuses on static graph models, which represent project tasks and dependencies at a single point in time. However, projects are dynamic in nature, with tasks, resources, and dependencies constantly evolving over the project's lifecycle. There is a need for more research on dynamic graph models that can capture the changes in project structures and relationships over time, providing a more accurate representation of project progression and enabling better decision-making and adaptability in project management (Qi et al. 2021; Radini et al. 2023; Stilla & Xu 2023).

Although graph theory has shown promise as a valuable tool in project management, there are several limitations in the existing research that need to be addressed. More empirical studies are needed to validate and improve graph-based models, and standardization efforts should be undertaken to facilitate comparison and synthesis of findings. Additionally, researchers should focus on developing models that better represent the complexities of real-world projects, including task interdependencies, uncertainty, and resource constraints. Finally, more research on dynamic graph models is necessary to account for the evolving nature of projects. Addressing these limitations will enhance the practical applicability and effectiveness of graph theory in project management, ultimately contributing to improved project outcomes (Debrah, Chan & Darko 2022; Roxas et al. 2023).

In my opinion, while graph theory offers numerous benefits for project management, such as visualizing project components and supporting decision-making, it also presents notable challenges. The complexity of creating and analyzing graphs, coupled with the need for specialized expertise, can be daunting for project managers. Moreover, the accuracy of graph representations hinges on the quality of input data, and interpreting these graphs may involve subjective judgment. Communicating the implications of graph theory to stakeholders can also pose difficulties. Despite these challenges, the potential of graph theory in project management remains significant. However, addressing the limitations outlined, such as the lack of empirical studies, standardization issues, oversimplified models, and the absence of research on dynamic graphs, is crucial for maximizing its effectiveness. By acknowledging and tackling these challenges, the practical utility of graph theory in project management can be enhanced, leading to improved project outcomes.

2.11. Agile project management

Agile project management is a modern approach to managing projects that has become increasingly popular over the last few decades. The concept of agile project management was first introduced in the 1990s, and it has since evolved into a widely used methodology for managing complex projects. In this history of agile project management, I will explore the evolution of the methodology and the key concepts that have made it so successful (Gemino, Horner Reich & Serrador 2020; Mergel, Ganapati & Whitford 2021).

The history of agile can be traced back to the mid-1990s when a group of software developers came together to discuss ways to improve the traditional Waterfall model of software development. This group, which included Kent Beck, Ward Cunningham, and Ron Jeffries, came up with the Agile Manifesto in 2001. The Agile Manifesto is a set of guiding principles for software development that emphasizes customer satisfaction, working software, and collaboration (Ciric et al. 2019; Stoddard, Gillis & Cohn 2019).

Since then, the agile methodology has evolved and grown beyond software development, becoming a popular approach to project management in many industries. The evolution of agile has led to the development of numerous frameworks, such as Scrum, Kanban, and Lean, which are tailored to specific industries and project types (Breyter 2022; Wafa et al. 2022).

Out of these discussions, the Agile Manifesto was born. The manifesto was a set of guiding principles for software development that emphasized collaboration, flexibility, and adaptability. The manifesto was grounded on the concept that the development of software is a cyclical process necessitating continual feedback and adjustment (Chandrachooodan, Radhika & Palappan 2023; Savio 2022).

The Agile Manifesto was a defining moment in the history of agile project management. The manifesto emphasized four key values:

1. Individuals and interactions over processes and tools
2. Working software over comprehensive documentation
3. Customer collaboration over contract negotiation
4. Responding to change over following a plan

These values became the foundation of agile project management, and they continue to guide the methodology to this day.

Agile methodologies encompass a collection of principles and practices that emphasize adaptability, teamwork, and ongoing enhancement in project management. Originally devised for software development projects, Agile methodologies have since been extended to a broad spectrum of industries and projects (Irtaymah, Shaari & Ahmed 2022; Malla & Prasad 2022).

In my opinion, the evolution and widespread adoption of agile project management represent a significant shift in how projects are approached and executed. The Agile Manifesto, born out of a necessity to improve upon traditional methods like the Waterfall model, highlights the importance of collaboration, flexibility, and adaptability. Personally, I find these values resonate deeply with the dynamic nature of modern project environments, where requirements and circumstances can change rapidly. Moreover, the emphasis on individuals, working software, customer collaboration, and responsiveness to change aligns well with my own experiences in project management. I believe that the principles and practices of agile methodologies offer a pragmatic and effective framework for navigating the complexities of project management across various industries.

2.11.1. Agile methodologies

Let's take a look at some of the widely adopted Agile approaches

1. Scrum

Scrum is an Agile methodology that emphasizes teamwork, collaboration, and iterative development. It is based on the idea of sprints, which are short, focused periods of development that typically last two to four weeks. During a sprint, the team focuses on delivering a specific set of features or functionality. Scrum also advocates for frequent meetings, including daily check-ins, sprint planning, and retrospectives, to guarantee alignment and progress within the team (Esteki, Gandomani & Farsani 2020; Ozkan et al. 2022).

2. Kanban

Kanban is an Agile methodology that focuses on visualizing and managing work in progress. It is based on the idea of a kanban board, which is a visual representation of the work that needs to be done. The board typically has columns that represent different stages of the work, such as "to do," "in progress," and "done." Each task or item is represented by a card or sticky note, which is moved across the board as it progresses through the stages. Kanban emphasizes limiting work in progress, so that the team can focus on completing tasks before starting new ones (Ozkan et al. 2022; Zayat & Senvar 2020).

3. Lean

Lean is an Agile methodology that is based on the principles of lean manufacturing. It emphasizes the elimination of waste, continuous improvement, and customer focus. Lean focuses on delivering value to the customer as quickly as possible, by identifying and eliminating unnecessary steps in the development process. Lean also emphasizes the importance of feedback and learning, so that the team can continuously improve their processes and deliver better results (Albuquerque, Torres & Berssaneti 2020; Cusumano et al. 2021; Malla & Prasad 2022).

4. Extreme Programming (XP)

Extreme Programming is an Agile methodology that emphasizes the importance of software engineering practices such as continuous integration, automated testing, and pair programming. XP focuses on delivering high-quality software that meets the customer's needs. XP also emphasizes the importance of frequent communication and collaboration between team members, as well as regular feedback and continuous improvement (Alsharari et al. 2023; Imran & Soomro 2022; Zasa, Patrucco & Pellizzoni 2021).

In my opinion, the Agile methodologies outlined, including Scrum, Kanban, Lean, and Extreme Programming (XP), offer valuable frameworks for software development teams seeking to enhance collaboration, efficiency, and product quality. Each methodology presents unique approaches to project management, emphasizing aspects such as iterative development, visualizing work, eliminating waste, and fostering software engineering practices. While all these methodologies share common principles such as teamwork, customer focus, and continuous improvement, they also cater to different team dynamics and project requirements. Personally, I find the emphasis on regular feedback, continuous learning, and adaptability inherent in Agile methodologies particularly compelling, as they align well with the dynamic nature of modern software development environments.

2.11.2. Agile methodologies principles

Agile methodologies share several key principles, including:

1. Customer collaboration:

Agile methodologies prioritize collaboration with the customer, to ensure that the team is delivering what the customer needs (Alsharari et al. 2023).

2. Iterative progression:

Agile approaches stress the importance of delivering functional software in regular, small increments instead of postponing for a comprehensive, final product (Martin 2023).

3. Flexibility:

Agile approaches place a high value on adaptability and flexibility, enabling the team to react effectively to evolving requirements and market dynamics (Almanaseer et al. 2022).

4. Teamwork:

Agile methodologies prioritize teamwork and collaboration, to ensure that everyone is working together towards a common goal (Papadakis & Tsironis 2020).

5. Continuous improvement:

Agile methodologies prioritize continuous learning and improvement, so that the team can deliver better results over time (Malla & Prasad 2022).

These are just a few of the many agile methodologies that have emerged over the years. Each methodology has its own strengths and weaknesses, and different methodologies may be better suited to different types of projects.

In my opinion, agile methodologies represent a modern approach to project management that prioritizes customer collaboration, iterative progression, flexibility, teamwork, and continuous improvement. These principles underscore the strengths of agile methodologies in adapting to evolving requirements and delivering value to stakeholders.

2.11.3. Agile project management benefits

Agile project management has become increasingly popular over the years due to its many benefits. Here are some of the key benefits of Agile project management:

1. Improved Collaboration

Agile project management underscores the significance of team collaboration and unity, which can foster improved communication and a more efficient work setting. By incorporating all team members in the decision-making procedure and promoting transparent communication, Agile project management can aid in establishing trust and instilling a sense of responsibility and ownership (Mergel, Ganapati & Whitford 2021).

2. Increased Flexibility

Agile project management is designed to be flexible and adaptable, which allows teams to respond to changing requirements and market conditions. Agile project management provides teams with the flexibility to modify their strategies and priorities as required, which can aid in averting setbacks and maintaining the project's trajectory (Stoddard, Gillis & Cohn 2019).

3. Faster Delivery

Agile project management emphasizes iterative development, which allows teams to deliver working software more quickly than traditional project management methods. By segmenting the project into smaller, manageable parts, Agile project management can assist in ensuring that advancement is occurring and that the project is progressing (Breyter 2022; Wafa et al. 2022).

4. Better Quality

Agile project management emphasizes the importance of software engineering practices such as automated testing and continuous integration, which can lead to better quality software. By focusing on delivering working software in small increments and continuously testing and refining it, Agile project management can help ensure that the final product meets the customer's needs and expectations (Chandrachooodan, Radhika & Palappan 2023; Savio 2022).

5. Enhanced Customer Satisfaction

Agile project management emphasizes customer collaboration, which can lead to a better understanding of customer needs and higher levels of customer satisfaction. By involving the customer in the development process and getting regular feedback, Agile project management can help ensure that the final product meets the customer's expectations and requirements (Malla & Prasad 2022).

6. Better Risk Management

Agile project management involves regular reviews and assessments of project progress, which can help identify and mitigate potential risks. By identifying and addressing risks early on, Agile project management can help prevent costly delays and ensure that the project stays on track (Irtaymah, Shaari & Ahmed 2022).

7. Greater Transparency

Agile project management highlights the importance of transparency and candid communication, which can foster trust and ensure alignment among team members. By making project details and advancements accessible to all team members, Agile project management can help guarantee that everyone is striving towards common objectives and that there are no unexpected developments (Leong et al. 2023).

8. Improved Time-to-Market

Agile project management can help teams deliver products to market more quickly than traditional project management methods. By breaking the project down into smaller, manageable chunks and focusing on delivering working software in small increments, Agile project management can help ensure that the product is delivered to market as quickly as possible (Almanaseer et al. 2022; Marnada et al. 2022).

9. Increased Adaptability

Agile project management is structured to be malleable and adjustable, enabling teams to react to evolving requirements and market trends. By concentrating on delivering operational software in small portions and modifying strategies and priorities as necessary, Agile project management can assist in ensuring that the project remains on course and that the end product aligns with the client's expectations (Alsharari et al. 2023; Martin 2023).

In my opinion, Agile project management stands out as a transformative approach that offers a plethora of advantages for modern project teams. The emphasis on collaboration not only enhances communication but also nurtures a culture of trust and shared responsibility among team members. Flexibility is another key feature that allows teams to swiftly adapt to changing circumstances, thereby mitigating potential setbacks and ensuring the project's success. Moreover, the iterative nature of Agile promotes faster delivery of high-quality software, leading to heightened customer satisfaction. Through regular reviews and transparent communication, Agile project management also aids in effective risk management and fosters greater transparency within the team. Overall, Agile methodologies not only expedite time-to-market but also empower teams to remain adaptable and responsive in today's dynamic business landscape.

2.11.4. Agile project management challenges

While Agile project management has gained popularity over the years due to its many benefits, it also presents a set of unique challenges. Here are some of the key challenges of Agile project management:

1. Resistance to Change

Implementing Agile project management requires a significant shift in mindset and culture. Many organizations are resistant to change, and it can be difficult to get everyone on board with the new way of doing things. This resistance can come from management, team members, or stakeholders who are comfortable with traditional project management methodologies (Alsharari et al. 2023).

2. Lack of Experience

Agile project management is a relatively new approach to project management, and many organizations lack the experience and expertise needed to implement it effectively. Teams may need additional training and coaching to understand the principles and practices of Agile project management (Imran & Soomro 2022).

3. Managing Scope

Agile project management emphasizes flexibility and adaptability, but this can make it challenging to manage scope. Without a clear plan and roadmap, it can be difficult to define the boundaries of the project and ensure that it stays on track (Kaushik & Datta 2020).

4. Balancing Priorities

Agile project management involves continuous prioritization and re-prioritization of tasks and requirements. This can be challenging when stakeholders have different priorities or when there are competing demands on the team's time and resources (Imran & Soomro 2022).

5. Lack of Documentation

Agile project management prioritizes operational software over extensive documentation. This method can result in quicker delivery and increased adaptability, but it may also pose challenges in monitoring progress and ensuring team alignment (Marnada et al. 2022; Martin 2023).

6. Distributed Teams

Agile project management is designed to work well with co-located teams, but it can be challenging when team members are distributed across different locations and time zones. Communication and collaboration can be more difficult, and it can be harder to build trust and maintain a cohesive team culture (Almanaseer et al. 2022).

7. Technical Debt

Agile project management emphasizes delivering working software quickly, but this can sometimes lead to technical debt. Technical debt refers to the accumulation of design or implementation flaws that can lead to problems down the line. It can be challenging to balance the need for speed with the need for quality and sustainability (Malla & Prasad 2022).

8. Scaling Agile

Agile project management is often most effective for small to medium-sized projects, but it can be challenging to scale up to larger projects or organizations. The larger the project or organization, the more complex the coordination and communication becomes (Gemino, Horner Reich & Serrador 2020).

9. Resistance from Stakeholders

Agile project management emphasizes customer collaboration, but this can be challenging when stakeholders have different expectations or are resistant to working in an Agile way. It can be difficult to build trust and maintain open communication with stakeholders who are used to traditional project management methods (Stoddard, Gillis & Cohn 2019).

10. Planning and Estimation

Agile project management is designed to be flexible, but it still requires some level of planning and estimation. This can be challenging when requirements are uncertain or when there are many unknowns. Teams may need to adjust their plans and estimates frequently to stay on track (Zayat & Senvar 2020).

In my opinion, the challenges outlined in the literature review highlight the complexities that organizations face when adopting Agile project management methodologies. While the benefits of Agile are well-documented, such as increased flexibility and faster delivery, it's crucial to acknowledge the hurdles that come with this transition. Resistance to change, lack of experience, and difficulties in managing scope and priorities underscore the need for careful planning and ongoing support during the Agile implementation process. Additionally, the issues surrounding documentation, distributed teams, technical debt, and scaling Agile further emphasize the importance of finding a balance between agility and stability within project management practices. Overcoming these challenges requires a concerted effort from all stakeholders involved, along with a willingness to adapt and iterate as necessary to ensure the success of Agile projects in diverse organizational contexts.

2.11.5. Comparison of agile frameworks

Research has been conducted on the effectiveness of different agile frameworks. One study by researchers at the University of Tartu in Estonia found that Scrum was more effective than traditional Waterfall methods for managing software development projects. The study also found that teams using Scrum had higher levels of communication and collaboration than teams using Waterfall methods (Esteki, Gandomani & Farsani 2020; Ozkan et al. 2022).

Another study by researchers at the University of Sydney found that Kanban was effective in managing software development projects with high levels of uncertainty and volatility. The research discovered that Kanban facilitated teams in swiftly adapting to alterations in project scope and specifications (Wafa et al. 2022; Zayat & Senvar 2020).

A study by researchers at the University of Groningen in the Netherlands found that organizations using Agile methodologies experienced higher levels of customer satisfaction and were better able to adapt to changing market conditions. The study also found that Agile methodologies resulted in faster time-to-market and improved team morale (Irtaymah, Shaari & Ahmed 2022; Malla & Prasad 2022).

A study by researchers at the University of Helsinki in Finland found that the Agile framework of Extreme Programming (XP) was effective in managing software development projects with high levels of uncertainty and complexity. The study found that XP allowed teams to deliver high-quality software quickly and efficiently, while also promoting collaboration and communication between team members (Papadakis & Tsironis 2020).

A study by researchers at the University of South Australia found that Agile methodologies were effective in managing construction projects. The study found that Agile methodologies allowed construction teams to respond quickly to changes in project scope and requirements, resulting in improved project outcomes (Gemino, Horner Reich & Serrador 2020).

A study by researchers at the University of Portsmouth in the UK found that Agile methodologies were effective in managing marketing projects. The study found that Agile methodologies allowed marketing teams to respond quickly to changes in market conditions, resulting in improved marketing outcomes (Ciric et al. 2019).

Overall, research has found that different Agile frameworks can be effective in managing projects in a variety of industries and contexts. However, it is important for organizations to choose the appropriate framework for their specific project and industry, and to ensure that team members are adequately trained and supported in using the chosen framework (Zayat & Senvar 2020).

In my opinion, the findings presented in the literature review underscore the versatility and adaptability of Agile methodologies across diverse industries and project contexts. It's evident that various Agile frameworks, such as Scrum, Kanban, Extreme Programming (XP), and others, offer valuable advantages in terms of communication, collaboration, adaptability, and overall project success. Personally, I find it compelling how these methodologies have been shown to not only enhance project management but also positively impact customer satisfaction, time-to-market, and team morale. However, as highlighted in the research, the selection of the most suitable Agile framework should be informed by careful consideration of the project's specific requirements and the organization's industry dynamics. Moreover, ensuring adequate training and support for team members is crucial for maximizing the benefits of Agile implementation.

2.11.6. Implementation of Agile Methodologies

There are numerous effective strategies that organizations can adopt when executing Agile methodologies. A crucial strategy is to initiate on a small scale and progressively expand the application of Agile practices. This can help organizations avoid overwhelming their team members with too many changes at once (Albuquerque, Torres & Berssaneti 2020; Kaushik & Datta 2020).

Another best practice is to provide adequate training and support for team members who are new to Agile methodologies. This can include training sessions, workshops, and coaching to help team members understand the principles and practices of Agile project management.

It is also important to establish clear goals and metrics for measuring success when implementing Agile methodologies. This can help organizations track progress and identify areas where improvements are needed. Additionally, organizations should encourage open communication and collaboration between team members, and should promote a culture of continuous improvement and learning.

Agile project management has also emerged as a response to the limitations of traditional project management methodologies such as Waterfall. Waterfall is a linear, sequential approach that is well-suited to projects with well-defined requirements and a stable scope. However, it is less effective for projects that are highly uncertain or that require frequent changes to the scope or requirements. Agile project management provides a more flexible and adaptable approach that is better suited to these types of projects (Arnold & Bashir 2020; Zasa, Patrucco & Pellizzoni 2021).

The emergence of Agile project management has also been driven by changes in the workforce. Today's workforce is more diverse, geographically dispersed, and digitally connected than ever before. Agile project management offers a structure for overseeing remote or virtual teams, and for fostering cooperation and communication among team members, irrespective of their geographical location (Marnada et al. 2022; Martin 2023).

Lastly, the rise of Agile project management can be linked to an increasing acknowledgment of the significance of client satisfaction and user experience. Agile methodologies underscore the importance of delivering value to the client and involving them in the development journey. This client-focused approach has gained prominence as organizations strive to set themselves apart in a fiercely competitive marketplace (Marle 2020; Papadakis & Tsironis 2020; Santos & de Carvalho 2022).

In my opinion, the literature provides a comprehensive understanding of the various effective strategies within Agile methodologies. It is evident that starting on a smaller scale and gradually scaling up, coupled with adequate training and support for team members, forms a solid foundation for successful Agile implementation. Moreover, the emphasis on establishing clear goals, fostering open communication, and embracing a culture of continuous improvement resonates with my own experiences in project management. Additionally, the comparison drawn between Agile and Waterfall methodologies highlights the necessity for adaptability in today's fast-paced and ever-changing environment. The discussion on Agile's compatibility with a diverse and digitally connected workforce also aligns with contemporary trends. Personally, I believe that Agile's focus on client satisfaction and user experience reflects a fundamental shift towards a more customer-centric approach, which is crucial for staying competitive in today's marketplace. Overall, the literature underscores the importance of flexibility, collaboration, and responsiveness, which I believe are essential pillars for successful project management in any organization.

3. MATERIALS AND METHODS

3.1. Data source

Project management is a complex process that involves planning, organizing, and controlling resources to achieve specific goals. It is essential to collect relevant data to conduct research in this area. Once the research question is defined, it is necessary to identify the sources of information that can provide valuable insights into project management.

There are several sources of information that might be relevant for a project management research study. These sources could include academic papers, reports, case studies, and other published materials related to project management. These sources can help researchers understand key concepts and theories related to project management and provide insights into best practices and successful strategies.

Collecting and filtering relevant data is a critical step in the research process. Upon pinpointing potential information sources, it is crucial to sift through the data using specific keywords pertinent to the research query. This aids in ensuring that the gathered data is relevant and beneficial for the study.

For example, if the research question is focused on the role of leadership in project management, the researcher might search for academic papers, reports, and case studies that use keywords such as "leadership," "project management," "team management," and so on. By filtering the data based on these keywords, the researcher can focus on finding relevant information that can help answer the research question.

3.1.1. Academic source

Academic sources can be defined as a body of work produced by experts in a particular field of study. These experts are often scholars, researchers, or academics who have conducted studies and experiments to test hypotheses related to their area of expertise. The work is then published in peer-reviewed journals or academic books, which are rigorously reviewed by other experts in the same field. The process ensures that the research meets high standards of quality and accuracy, making it a valuable resource for other researchers.

These sources provide several advantages that make them valuable for research purposes. Firstly, academic sources are reliable because they are based on rigorous research methods and adhere to scholarly standards. This ensures that the information presented in these sources is accurate and trustworthy (Gusenbauer & Haddaway 2020; Khatter et al. 2021).

Secondly, academic sources offer extensive coverage of topics, making them suitable for conducting comprehensive research. Scholars frequently conduct extensive literature reviews of previous research, and this information is often synthesized and presented in academic publications. Consequently, academic sources are an excellent starting point for gathering information about a particular topic (Evans et al. 2020; Johnson, Adkins & Chauvin 2020).

Thirdly, academic sources are often recognized as authoritative sources of information. Researchers and experts in various fields use academic sources to support their arguments and validate their findings. Therefore, referencing academic sources in research papers and essays enhances the credibility of the work (Kington et al. 2021; Kretser et al. 2019).

Despite the numerous advantages of academic sources, there are some drawbacks associated with their use. Firstly, academic sources can be challenging to access. Many of these sources require subscriptions or fees, making them inaccessible to people who cannot afford them. Furthermore, some academic sources are only available in specialized libraries, which may not be accessible to everyone (Liu, Tang & Lim 2023; Scharrer, Bromme & Stadtler 2021).

Secondly, academic sources can sometimes be overly technical and difficult to understand. The language used in these sources can be very technical and specialized, making it challenging for non-experts to comprehend. This can make it difficult for some researchers to use academic sources effectively (Goldsack et al. 2022; Seaton, Loch & Lugosi 2022).

Finally, academic sources can also be biased. Scholars have their own biases and perspectives, which can influence the information presented in academic sources. Therefore, it is essential to consider the author's perspective and potential biases when using academic sources (Elster 2023; Garzón, Pavón & Baldiris 2019; Scharrer, Bromme & Stadtler 2021).

Google Scholar, Web of Science, and Scopus are three of the most widely used academic search engines in the world. Each of these platforms has its unique features and benefits that make them popular among researchers, academics, and students. It is the largest scientific database, covering a broad range of disciplines and sources. It provides easy access to scholarly literature, including articles, books, theses, and conference papers. The biggest advantage of Google Scholar is its comprehensive coverage of scientific literature, which makes it an ideal choice for conducting general searches or exploring interdisciplinary topics.

However, the downside of Google Scholar is that it lacks the quality control measures and structured metadata found in other databases, such as Scopus and Web of Science. This can lead to issues with data accuracy, completeness, and consistency. Additionally, Google Scholar's ranking system is based on citation counts, which can be biased towards well-known authors or institutions and may not reflect the actual impact of the research (Bamber et al. 2020; Zhu & Liu 2020).

Scopus is a multidisciplinary scientific database that covers over 25,000 journals across all fields of science, technology, medicine, and social sciences. It provides rich metadata and advanced search capabilities, allowing researchers to filter results by author, affiliation, document type, and more. Scopus also offers citation analysis tools that help researchers track the impact of their publications and identify potential collaborators.

Compared to Google Scholar, Scopus offers higher quality data and more structured information, which makes it easier to analyze and compare research findings. However, Scopus has a smaller coverage than Google Scholar, meaning that some studies that are available in Google Scholar might not be found in Scopus. Additionally, Scopus only indexes journals that meet certain quality standards, which can limit the availability of some research results (Martín-Martín et al. 2021; Pranckutė 2021).

Web of Science is a leading platform for scientific research that provides access to a vast collection of scholarly publications and citations. It is a citation index that covers over 20,000 scientific journals in all disciplines. It offers robust search capabilities and advanced analytics tools, including citation analysis, co-citation analysis, and bibliometric mapping. Web of Science also provides access to conference proceedings, books, and patents, making it a comprehensive resource for research.

A primary benefit of Web of Science is its focus on data integrity and organized metadata, which guarantees that researchers can rely on it and allows them to scrutinize and contrast outcomes with greater precision (Bamber et al. 2020; Harzing 2019). Data quality refers to the accuracy, completeness, and reliability of the information available on a platform. This is particularly important when conducting bibliometric analyses, as the accuracy of citation data is critical for drawing meaningful conclusions about research impact (Martín-Martín et al. 2021; Zhu & Liu 2020).

Web of Science's commitment to data quality is evident in its rigorous selection process for indexed journals. The platform only indexes high-quality journals that meet certain criteria, including peer review, editorial quality, and citation analysis. This ensures that the publications included in Web of Science are reliable and of high quality, providing researchers with a trustworthy source of data for their analyses.

In addition to its selection process for indexed journals, Web of Science also employs a team of expert editors who carefully curate the metadata associated with each publication. This includes detailed information about authorship, affiliations, funding sources, and more. Through the provision of organized metadata, Web of Science empowers researchers to carry out intricate analyses of their own studies and those of others, discerning patterns and tendencies that might not be readily observable from the unprocessed citation data alone (Birkle et al. 2020; Moskalenko et al. 2022).

For example, researchers can use Web of Science to examine patterns in co-authorship across different institutions or countries, or to identify key funding sources for research in a particular field. These insights can help researchers develop new collaborations, identify potential funding sources, and uncover emerging areas of research interest.

Web of Science's commitment to data quality also extends to its citation analysis tools. The platform uses a proprietary algorithm to ensure that citation data is accurate and up-to-date, allowing researchers to track the impact of their work over time and compare it to that of other researchers in their field. Citation analysis is a powerful tool for evaluating research impact, as it provides a quantitative measure of how often a particular publication has been cited by other scholars (Singh et al. 2021; Xu et al. 2020).

By using Web of Science's citation analysis tools, researchers can identify the most highly cited publications in their field, track changes in citation patterns over time, and even predict future research trends. This information can be particularly valuable for researchers seeking funding or tenure, as it provides evidence of the impact and significance of their research. In addition to its focus on data quality, Web of Science is also known for its user-friendly interface and powerful search capabilities (Alryalat, Malkawi & Momani 2019; Pranckutė 2021).

The platform allows users to easily search for publications by author, title, keyword, or other criteria, making it easy to find relevant research articles. Users can also save searches and set up alerts to receive notifications when new publications are added to the database that match their search criteria. Web of Science's search capabilities are enhanced by its use of advanced algorithms and machine learning models. These utilities allow the platform to swiftly scrutinize extensive data volumes and offer customized suggestions to users, taking into account their search history and predilections. This makes it easier for researchers to discover new publications and stay up-to-date with the latest research in their field (Monaghesh & Hajizadeh 2020; Zhu & Liu 2020).

Another strength of Web of Science is its ability to support cross-disciplinary research. The platform includes publications from a wide range of disciplines, including science, social science, arts, and humanities. This enables researchers to explore interdisciplinary connections and identify new research opportunities that cut across traditional disciplinary boundaries. For example, a researcher studying the intersection of biology and computer science could use Web of Science to identify publications that bridge these two fields, such as articles on computational biology or bioinformatics. By drawing on insights from multiple disciplines, researchers can develop innovative solutions to complex problems and make breakthrough discoveries that would not be possible within a single discipline (Harzing 2019; Monaghesh & Hajizadeh 2020).

Web of Science's cross-disciplinary capabilities are further enhanced by its integration with other research tools and databases For instance, the platform can be utilized in tandem with additional resources such as Scopus or PubMed to offer a more exhaustive perspective of the academic field. This integration enables researchers to access a wider range of publications and citations, helping them gain a more holistic and nuanced understanding of their research area.

Overall, Web of Science is a valuable tool for researchers seeking to conduct bibliometric analyses and evaluate the impact of their work. By providing access to high-quality data and structured metadata, the platform enables researchers to make informed decisions about their research and gain insights into the broader landscape of scholarly communication. Its user-friendly interface, powerful search capabilities, and cross-disciplinary capabilities make it a valuable resource for researchers in a wide range of fields (Bamber et al. 2020; Visser, van Eck & Waltman 2021; Zhu & Liu 2020).

Overall, the combination of data quality and structured metadata makes Web of Science a powerful tool for analyzing and comparing research results more accurately. By providing consistent and reliable information on authors, journals, and research topics, the platform enables researchers to make meaningful contributions to their respective areas of study. Therefore, due to the explained reasons the Web of Science used as academic source (Birkle et al. 2020; Moskalenko et al. 2022; Pranckutė 2021).

3.1.2. Non-academic source

Traditionally, research has focused on using academic sources such as journals, books, and articles. However, non-academic sources can also provide valuable information that can enrich a study. Non-academic sources can provide perspectives that are not found in academic literature, and they can help researchers to explore topics from a variety of angles. Non-academic sources refer to sources such as newspapers, magazines, blogs, and social media platforms that are not peer-reviewed or written by scholars. These sources include books, magazines, newspapers, government reports, personal interviews, historical documents, blogs, podcasts, videos, and social media posts. While these sources may not have undergone the same rigorous peer-review process as academic sources, they can still provide valuable insights into a research topic. These sources also provide several advantages that make them valuable for research purposes (De Bernardi et al. 2021; Kircaburun et al. 2021).

Firstly, non-academic sources are often more accessible than academic sources. Many of these sources can be accessed online for free, making them easily available to anyone with an internet connection. This makes non-academic sources ideal for conducting research on current events and trends. Secondly, non-academic sources can provide a broader range of perspectives. Unlike academic sources, non-academic sources are not limited to scholarly analysis and can include the opinions and experiences of individuals from different backgrounds and fields. This can enrich the data collected and provide a more holistic view of the topic being researched. Thirdly, non-academic sources can provide insights into the practical applications of research findings. For example, news articles about scientific discoveries often discuss how they may impact people's lives, which can be useful to researchers who want to understand the real-world implications of their studies (Alvesson & Sandberg n.d.; Rampton, Maybin & Roberts 2015).

Non-academic sources also have disadvantages that should be considered when using them for research purposes. Firstly, non-academic sources can lack credibility. Unlike academic sources, non-academic sources do not undergo rigorous peer-review processes, which can make their accuracy questionable. Therefore, it is essential to verify non-academic sources before using them for research. Secondly, non-academic sources can be biased. News outlets and social media platforms frequently have their own agendas and biases that can influence the information presented. Therefore, it is essential to consider the source's reputation and potential biases when using non-academic sources (Dahl 2009; Mendonça, Pereira & Ferreira 2018).

Finally, non-academic sources can also lack depth and detail. Unlike academic sources, which are often based on extensive research, non-academic sources may provide only a superficial understanding of a topic. Therefore, it is essential to supplement non-academic sources with academic sources to ensure a comprehensive understanding of the subject matter (De Bernardi et al. 2021).

Different sources exit to get the desire non-academic sources. One way of collecting non-academic sources for research is to use online databases. Many online databases provide access to non-academic sources such as newspapers, magazines, and government reports. For example, ProQuest provides access to over 90,000 newspapers and magazines from around the world. Similarly, JSTOR offers access to thousands of historical documents, government reports, and primary sources. These databases can be accessed through institutional libraries (Baltaru 2018; Pascarella & Terenzlni 2014)18; Pascarella & Terenzlni 2014).

Another way of collecting non-academic sources for research is to conduct personal interviews. Personal interviews can provide valuable insights into a research topic and can offer perspectives that may not be found in academic sources. To conduct personal interviews, researchers can reach out to experts in their field, community leaders, or individuals with first-hand experience of the topic being studied. Researchers can also use online platforms such as Zoom or Skype to conduct virtual interviews with subjects who are located far away.

In addition to personal interviews, researchers can also collect non-academic sources from archives and museums. Archives and museums hold vast collections of historical documents, photographs, artifacts, and other non-academic materials that can be used for research purposes. To access these sources, researchers can visit archives and museums in person or use online portals to search for relevant materials (Kircaburun et al. 2021; McIntyre et al. 2018).

An additional beneficial reservoir of non-academic data is social media. Platforms like Twitter, Facebook, Instagram, and LinkedIn offer a plethora of information on diverse subjects. Researchers can leverage social media to collect data, execute surveys, and scrutinize content generated by users. Nevertheless, it's crucial to remember that social media can also be a conduit for misinformation, hence researchers should exercise caution when utilizing social media as a source for research.

Blogs and podcasts are two other sources of non-academic information that can be useful for research purposes. Many bloggers and podcasters are experts in their respective fields and can provide valuable insights into a topic. Researchers can follow relevant blogs and podcasts and use them as sources of information. They can also reach out to bloggers and podcasters for interviews or to request additional information (De Bernardi et al. 2021; Rampton, Maybin & Roberts 2015).

Personal experience is another source of non-academic information that can be used for research purposes. Researchers can draw on their own experiences or the experiences of others to gain insights into a research topic. For example, a researcher studying the effects of climate change on farming can draw on the experiences of farmers who have been impacted by changing weather patterns.

Non-academic sources can also be found in popular culture. Movies, TV shows, music, and other forms of popular culture can reflect societal attitudes and beliefs about a particular topic. Researchers can analyze popular culture to gain insights into how a topic is perceived by the general public. However, it is important to note that popular culture may not always be an accurate reflection of reality. In addition to the sources mentioned above, researchers can also collect non-academic sources from government agencies and NGOs. Government reports, policy briefs, and other documents can provide valuable information on various topics. Similarly, NGOs often publish reports and articles on their areas of expertise that can be used for research purposes (Dahl 2009; Kircaburun et al. 2021).

When collecting non-academic sources for research, it is important to critically evaluate the sources. Non-academic sources may not have undergone the same rigorous peer-review process as academic sources, so it is important to assess their credibility and reliability. Researchers should consider factors such as the author's credentials, the publication date, and whether the information presented is supported by other sources (Merga & Mason 2020; Olmos-Peñuela, Castro-Martínez & D'Este 2014).

The processing of non-academic sources in research and business operations is a common practice in today's digital age. However, it is essential to process these sources ethically and legally, keeping in mind copyright laws and data privacy regulations. Processing non-academic sources requires careful consideration of the rights of writers and compliance with GDPR regulations (Herndl 1993; Olmos-Peñuela, Castro-Martínez & D'Este 2014).

One of the primary considerations when processing non-academic sources is copyright protection. Copyright laws protect the creative works of authors, including written works, images, and other original creations. Individuals and organizations must obtain permission from the copyright holder before using any copyrighted material. Unauthorized use of copyrighted material can result in legal action and negative consequences for individuals and organizations. Therefore, it is crucial to obtain permission from the copyright holder before using any copyrighted material (Adler et al. 2010; Gunn & Mintrom 2016; Xiao 2023).

There are several methods available to obtain permission to use copyrighted material. One way is to contact the writer directly and request permission to use their work. This approach is often used when working with individual writers or small businesses. For larger organizations, a licensing agreement may be necessary to obtain permission to use copyrighted material. Licensing agreements outline the terms and conditions under which the copyrighted material can be used, including the duration of use and compensation for the writer (Adler et al. 2010).

Open access and public domain resources provide free access to works that are not subject to copyright restrictions. Public domain works include those whose copyright has expired or those that were never protected by copyright law. Open access resources require attribution or a link back to the original source but do not require explicit permission.

Citing sources properly is also critical when processing non-academic sources. Proper citation not only gives credit to the original author but also demonstrates academic integrity. Adequate citation includes acknowledging the author's name, the title of the work, the date of publication, and the source of the work. Failure to cite sources correctly can lead to plagiarism and ethical concerns.

An additional factor to consider is the regulations surrounding data privacy. The General Data Protection Regulation (GDPR) presides over data protection and privacy for residents of the European Union (EU). This regulation stipulates stringent conditions for the collection, storage, and processing of personal data, encompassing details like names, email addresses, and other identifying particulars. Organizations are required to secure consent from individuals prior to gathering their personal data and must safeguard this data from unauthorized access and utilization (Baraliuc, Depreeuw & Gutwirth 2013; Xiao 2023).

In order to adhere to the GDPR, organizations are required to put in place suitable security protocols to safeguard personal data. This encompasses encrypting confidential data, limiting access to only authorized personnel, and guaranteeing that data is housed on secure servers. Furthermore, organizations are obligated to grant individuals the right to access and manage their personal data, inclusive of the right to demand the deletion of their data (Alvesson & Sandberg n.d.; De Bernardi et al. 2021).

Transparency is indeed a fundamental tenet of GDPR compliance. It is imperative for organizations to be forthright about the nature of the data they accumulate, the manner in which it is utilized, and the entities with whom it is shared. This means providing individuals with clear and concise information about the processing of their personal data and obtaining their explicit consent before processing it. Fair use allows for the use of copyrighted material without permission under certain circumstances. The criteria for fair use include using the work for educational or research purposes, commenting on or criticizing it, or transforming it into a new work. Fair use is determined on a case-by-case basis and requires a careful analysis of the specific circumstances surrounding the use of copyrighted material (Bravo et al. 2019; Colombo & Piva 2012).

When determining whether the use of copyrighted material meets the criteria for fair use, several factors are considered. These include the purpose and character of the use, the nature of the copyrighted work, the amount of the work used, and the effect of the use on the market value of the original work. If the use of copyrighted material does not meet the criteria for fair use, obtaining permission from the copyright holder is necessary (Kircaburun et al. 2021; Mendonça, Pereira & Ferreira 2018).

Ethical concerns when processing non-academic sources are also crucial to consider. Issues such as bias, accuracy, and reliability are common in non-academic sources. Non-academic sources may contain subjective viewpoints that can influence the interpretation of data. Therefore, it is essential to verify the accuracy and reliability of the sources used in research. Fact-checking and cross-referencing with other sources can help mitigate these concerns.

Another ethical concern is the use of personal data obtained from non-academic sources. Indeed, individuals possess the right to privacy and the safeguarding of their personal data. It is incumbent upon organizations to ensure that an individual's personal data is protected and utilized solely for the objectives for which it was gathered. Failure to do so can result in legal action, negative publicity, and damage to an organization's reputation.

Fig. 3 depicts the steps and tools utilized for data collection, specifically tailored for non-academic data. Additionally, Fig. 4 outlines the procedures for obtaining the desired dataset through web scraping.

Figure 3. Data collection tools and steps (own research)

Figure 4. Web data collection from search to scrape steps (own research)

3.2. Application of graph theory

After the pertinent data has been gathered and sifted based on specific keywords tied to the research query, you can commence the application of graph theory to construct networks. Graph theory is a mathematical structure frequently employed to model intricate systems, such as social networks, biological networks, and information networks. By applying graph theory to project management data, I can create visual representations of the relationships between different concepts and keywords.

There are different approaches to creating networks using graph theory. One common approach is to use the co-occurrence of keywords as the basis for creating edges between nodes in the network. In this approach, if two keywords appear together frequently in the dataset, an edge is created between them in the network. The more often the two keywords appear together, the stronger the edge between them (Giabbanelli & Tawfik 2020; Wu et al. 2023).

Another approach is to use the frequency of keyword occurrence as the basis for determining node size or edge weight. In this approach, keywords that appear more frequently in the dataset are represented by larger nodes in the network. Similarly, edges between nodes that represent frequently occurring keywords are given more weight in the network.

Creating networks based on co-occurrence or frequency of keyword occurrence can help us identify important concepts and relationships in the data. For example, if there are frequent co-occurrences between the keyword's "risk" and "project management," it may suggest that risk management is an important consideration in project management. Similarly, if certain keywords have high node sizes or edge weights, it may indicate that these concepts are particularly relevant or significant in the dataset.

Once the networks have been created, I can visualize and analyze the data to gain insights into project management practices and strategies. For example, I can identify clusters of keywords that are closely related and examine their relationships to other clusters. I can also measure the centrality of different nodes or groups of nodes, which can provide information about which concepts or themes are most central to the dataset. By scrutinizing the networks, I can attain a more profound comprehension of the interconnections between various concepts and how these might influence the results of project management. This approach can provide valuable insights that can guide decision-making and strategy in project management (Javaid et al. 2022; Nandwani & Verma 2021).

In addition to creating networks based on keyword co-occurrence or frequency of occurrence, there are other approaches that I can use to create networks in project management. For example, I can use semantic analysis to identify relationships between keywords based on their meanings, rather than their co-occurrence or frequency. I can also use network clustering algorithms to identify subgroups of nodes that are highly connected within the network.

Regardless of the specific approach used, creating networks in project management research can provide valuable insights into the relationships between different concepts and keywords. By using graph theory to model project management data, I can gain a deeper understanding of the factors that contribute to successful project outcomes. I can also use this information to develop recommendations for improving project management practices and strategies (Arosio, Martina & Figueiredo 2020; Tanglay et al. 2023).

The process of creating networks can be a complex and challenging task, particularly when attempting to represent a large number of nodes and their relationships accurately. This is where software applications like VOS or other third-party tools come in handy, as I provide users with the ability to create visual representations of networks quickly and efficiently.

However, it's worth noting that not everyone has access to such software or may find it difficult to use. In these cases, drawing networks manually can be an alternative solution. This may involve sketching out the network on paper first before transferring it to a digital format for further editing. Regardless of the method used, it's important to ensure that the resulting drawing accurately represents the relationships between different nodes in the network. This means taking the time to study the data and understand how the nodes are connected, so that the drawing can clearly illustrate these connections (Arosio, Martina & Figueiredo 2020; Murphy et al. 2022).

One way to achieve this is by using different shapes or symbols to represent different types of nodes. For example, circles might be used to represent people, while squares could represent organizations. Arrows can also be used to indicate the direction of the relationship between nodes, such as whether one node is influencing another. Another consideration when drawing networks is the layout of the nodes and their connections. A well-designed layout can make it easier for viewers to understand the relationships between nodes and identify key connections or patterns. There are several different approaches to network layout, including hierarchical, circular, and radial layouts (Giabbanelli & Tawfik 2020; Wu et al. 2023).

Hierarchical layouts are useful when there is a clear hierarchy within the network, with some nodes having more influence or importance than others. In this type of layout, nodes are arranged in levels, with each level representing a different degree of influence or importance. Circular and radial layouts are useful for networks where there isn't a clear hierarchy, as I allow all nodes to be displayed at the same level of importance.

The size and intricacy of the network are crucial factors to consider when planning its layout. For exceptionally large networks, it might be beneficial to decompose the network into smaller sub-networks or employ zooming utilities to enable viewers to concentrate on particular sections of the network. Besides layout and design considerations, it’s essential to ensure that the diagram is visually engaging and easy to interpret. This can be accomplished by utilizing colors and shading to emphasize key nodes or links, as well as incorporating labels or annotations to offer supplementary context for the data. This approach can enhance the readability and comprehension of the network, making it a more effective tool for analysis and communication. Overall, creating accurate and effective network drawings requires careful planning and attention to detail. By taking the time to understand the data and selecting the right tools and techniques, users can create visual representations of networks that effectively communicate complex relationships and patterns (Murphy et al. 2022; Tanglay et al. 2023).

Analyzing networks is a potent method for deciphering the composition and behavior of intricate systems. Through the application of diverse algorithms, it’s possible to delve into the connections among various elements and pinpoint significant participants or groupings within the network. Regardless of whether the analysis is conducted manually or via automated techniques, it can yield meaningful understanding across a broad spectrum of phenomena, encompassing everything from social networks and biological systems to transportation infrastructures and financial markets. After the networks have been established, they can be scrutinized using an array of network analysis algorithms. These algorithms can shed light on the network’s architecture, highlight critical nodes or groupings, and uncover patterns of correlation among different notions. Network analysis algorithms come in various forms, each possessing its unique advantages and limitations (Sergiou et al. 2020; Vattai & Mályusz 2020).

Centrality analysis is a category of algorithms that assesses the significance of network nodes. Various metrics are used in this analysis, including degree centrality, which quantifies a node’s edges, betweenness centrality, which evaluates how often a node appears on the shortest paths among other nodes, and eigenvector centrality, which estimates a node’s influence considering the impact of its adjacent nodes. Another algorithm category is community detection, which organizes nodes into clusters based on their connection patterns. These algorithms can expose the network’s inherent structures and assist in pinpointing node subgroups that have denser connections with each other compared to the rest of the network (Shetty & Raghu 2022; Valeri & Baggio 2021).

Path analysis represents another category of algorithms, focusing on the exploration of paths among nodes within the network. These algorithms can uncover influence patterns between nodes and assist in determining causal associations among various concepts. The selection of an algorithm is contingent on the research query and the attributes of the network under scrutiny. Certain algorithms might be more fitting for compact, dense networks, whereas others might be more advantageous for extensive, sparse networks.

In addition to these automated methods, network analysis can also be performed manually. Manual methods involve visually inspecting the network, identifying important nodes and clusters, and analyzing the relationships between them. Manual methods can be time-consuming and subjective, but I can also provide a deeper understanding of the nuances of the network that may be difficult to capture with automated methods.

A key benefit of network analysis is its ability to uncover patterns and associations that might not be evident through other data analysis methods. For instance, discerning the most influential individuals in a social network could be challenging when solely considering their personal attributes. However, by examining the network’s structure, it might be feasible to pinpoint pivotal individuals who maintain numerous connections and hold central roles within the network (Ma, Zhou & Li 2021; Punjani et al. 2022).

Network analysis can also be used to identify bottlenecks or vulnerabilities in a system. For example, in a transportation network, network analysis can help identify key nodes or links that, if disrupted, could have a significant impact on the overall functioning of the system. Another application of network analysis is in the study of disease transmission. By modeling the spread of a disease through a network of contacts, I can develop strategies for controlling the spread of the disease and identifying individuals who are at high risk of infection (Ni, Lin & Shen 2019; Sarker et al. 2021).

Network analysis serves as a potent instrument for comprehending the structure and dynamics of intricate systems. Upon the creation and illustration of the networks, I can employ a variety of network analysis algorithms to delve into the inherent patterns and associations present in the data.

Centrality analysis is one algorithmic approach I can employ. It evaluates the significance of nodes in the network based on their connections or impact. There are multiple centrality metrics, such as degree centrality, which quantifies the count of edges linked to a node, betweenness centrality, which assesses how frequently a node appears on the shortest paths among other nodes, and eigenvector centrality, which estimates a node's influence considering the impact of its adjacent nodes.

Another type of algorithm is community detection, which groups nodes together into clusters based on their connectivity patterns. By analyzing these clusters, I can identify subgroups of nodes that are more densely connected to each other than to the rest of the network. This can help them gain insights into the underlying structure of the network and how different components are interrelated.

The selection of an algorithm is contingent on the research query and the attributes of the network under scrutiny. Certain algorithms might be more fitting for compact, dense networks, whereas others might be more advantageous for extensive, sparse networks.

In addition to automated methods, network analysis can also be performed manually. Manual methods involve visually inspecting the network, identifying important nodes and clusters, and analyzing the relationships between them. While manual methods can be time-consuming and subjective, they can also provide a deeper understanding of the nuances of the network that may be difficult to capture with automated methods (Britto et al. 2018; Nabki et al. 2017; Priyadarshini & Puri 2021).

A significant benefit of network analysis is its ability to disclose patterns and associations that might not be immediately evident through other forms of data analysis. For instance, within a social network, pinpointing the most influential individuals could be challenging when solely considering their personal traits. However, by scrutinizing the network's structure, it might be feasible to identify prominent individuals who maintain a multitude of connections and hold central roles within the network.

Network analysis can also be used to identify bottlenecks or vulnerabilities in a system. For example, in a transportation network, network analysis can help identify key nodes or links that, if disrupted, could have a significant impact on the overall functioning of the system.

Another application of network analysis is in the study of disease transmission. By modeling the spread of a disease through a network of contacts, I can develop strategies for controlling the spread of the disease and identifying individuals who are at high risk of infection.

In essence, network analysis is a potent instrument for deciphering the structure and dynamics of intricate systems. By employing a range of algorithms, I can delve into the relationships among various entities, pinpoint key individuals or groups within the network, and reveal patterns that might not be immediately evident through other forms of data analysis. Whether conducted manually or via automated methods, network analysis can yield valuable insights into a broad spectrum of phenomena, from social networks to biological systems, and from transportation systems to financial markets. However, the analysis of individual networks might not offer a comprehensive view of the data. One strategy that I can adopt to attain a more profound understanding of the data is to juxtapose different types of networks (Malyusz, Hajdu & Vattai 2021; Shetty & Raghu 2022; Valeri & Baggio 2021).

One way to compare networks is by examining the frequency of keyword occurrence. This approach involves identifying the most common keywords within the dataset and mapping out the connections between them. By creating a network based on keyword frequency, I can gain insights into the most important concepts and themes within the dataset. They can also identify how these concepts relate to each other and how they contribute to the overall structure of the network.

Another approach is to create networks based on co-occurrence time. This involves analyzing how often keywords appear together over time and mapping out the relationships between them. By creating a network based on co-occurrence time, I can gain insights into how different concepts relate to each other over time and how they evolve and change. This approach can be particularly useful when studying trends or changes over time.

I can also compare networks that focus on specific subtopics within project management. For example, they could analyze networks that focus on risk management, stakeholder management, or scheduling. By comparing these networks, I can identify the commonalities and differences between different subtopics and gain a deeper understanding of the key concepts and relationships within each one. This will enable them to identify unique opportunities for improvement that are specific to each subtopic.

In addition to comparing networks based on their content, I can also compare networks based on their structural properties. For example, they might compare networks based on their size, density, or centrality. By comparing these metrics across different networks, I can identify the key differences and similarities between them and gain insights into the structural properties that drive these differences.

Another way to compare networks is by using network metrics. Network metrics are quantitative measures that describe different aspects of the network, such as its size, density, or centrality. By comparing these metrics across different networks, I can identify the key differences and similarities between them. For example, they might compare the degree centrality of different nodes in the network to identify the most important players in each network.

Through the comparison of diverse network types, I can attain a more profound understanding of the inherent patterns and associations in the data. These insights can serve to guide subsequent research, formulate new conjectures, or fine-tune existing theories. Additionally, comparisons between different networks can reveal unique opportunities for improvement that would not have been apparent through an analysis of individual networks.

One potential application of network analysis in project management is to create networks based on feedback from stakeholders. By analyzing the connections between stakeholders and the topics they discuss, I can identify areas where there is strong agreement or disagreement. This information can be used to guide project decisions and improve stakeholder engagement.

An additional potential utilization of network analysis lies in the realm of marketing research. By scrutinizing the links between consumers and brands within a network, I can acquire insights into the social dynamics that steer consumer behavior. This knowledge can aid companies in crafting more impactful marketing strategies that align with their target audience's preferences.

3.3. Using NLP in graph theory

Indeed, graph theory and network science are potent methodologies for deciphering the structure and dynamics of intricate systems. These disciplines involve the examination of networks, which are composed of nodes and edges symbolizing entities and their interconnections. One domain where graph theory and network science have proven particularly beneficial is in Natural Language Processing (NLP). NLP is a branch of artificial intelligence that concentrates on empowering computers to comprehend, interpret, and generate human language. NLP has a multitude of applications, ranging from chatbots and virtual assistants to sentiment analysis and machine translation.

Nonetheless, a challenge in NLP is converting unstructured text into structured data that can be analyzed using graph theory and network science. One strategy to tackle this challenge is to employ NLP techniques to extract pertinent information from text and depict it as a network. For instance, researchers might utilize Named Entity Recognition (NER) to pinpoint key entities within a text, such as individuals, organizations, or locations. They might then construct a network where each node signifies an entity and the edges denote relationships between them. As in this study I used the NLP to obtain the keywords in sources (Karhadkar et al. 2022; Pais, Cordeiro & Jamil 2022).

Another approach to using NLP in graph theory and network science is to focus on the relationships between words and concepts within a text. For example, researchers might use co-occurrence analysis to identify which words tend to appear together in a text. They might then create a network where each node represents a word, and the edges represent co-occurrence between words.

Once the network is created, researchers can apply various network analysis algorithms to gain insights into the structure and dynamics of the text. For example, they might use clustering algorithms to group words or entities together based on their similarity. Alternatively, they might use centrality algorithms to identify the most important words or entities within the text(Sarker 2021a; Sarker et al. 2020).

One way that NLP can benefit graph theory and network science is by providing more granular data for analysis. For example, researchers might use sentiment analysis to assign positive or negative values to each node in the network. This can help identify which entities or concepts are associated with positive or negative sentiment and how this sentiment changes over time (Dogra et al. 2022; Karhadkar et al. 2022; Sarker 2021a, 2022).

Another benefit of using NLP in graph theory and network science is that it can enable more sophisticated analysis of text data. For example, researchers might use topic modeling to identify the key topics within a text and map out the relationships between them. They might then create a network where each node represents a topic, and the edges represent the relationships between them. This approach can provide valuable insights into the underlying structure of the text and how different topics relate to each other (Moselhi, Bardareh & Zhu 2020; Ridzuan & Wan Zainon 2019; Sarker et al. 2021).

NLP can also be used to identify key phrases or keywords within a text that are relevant to a particular research question. By extracting these phrases or keywords and mapping out the connections between them, researchers can gain insights into the most important concepts or themes within the text.

One potential application of using NLP in graph theory and network science is in the analysis of social media data. By creating networks based on social media interactions, researchers can gain insights into the social dynamics that drive behavior on these platforms. For example, they might analyze the connections between users and the topics they discuss to identify areas of agreement or disagreement (Kendzierskyj et al. 2019; Nadikattu 2020).

Another potential application is in the field of finance. By analyzing news articles and other textual data related to financial markets, researchers can gain insights into the factors that drive market trends and fluctuations. By creating networks based on the relationships between key concepts and entities within the data, researchers can identify patterns and relationships that may not be apparent from other types of analysis (Lähnemann et al. 2020; Saura 2021).

3.4. Suggestion method to project manager

One challenge in agile project management is how to effectively analyze and manage project data in real-time. Graph theory and natural language processing (NLP) can be used to address this challenge by enabling project managers to gain insights into the structure and dynamics of their projects.

One suggestion method for project managers is to use graph theory and NLP to identify patterns and trends within project data. For example, project managers might use NLP to analyze project reports and identify key concepts or themes that are relevant to the project. They might then create a network where each node represents a concept or theme, and the edges represent relationships between them.

Once the network is created, project managers can apply various network analysis algorithms to gain insights into the structure and dynamics of the project. For example, they might use centrality algorithms to identify the most important concepts or themes within the network. Alternatively, they might use clustering algorithms to group related concepts or themes together based on their similarity.

An additional method suggested for project managers is to employ graph theory and NLP to pinpoint bottlenecks or vulnerabilities in the project. For instance, project managers might utilize graph theory to scrutinize the dependencies among tasks in the project and identify which tasks are critical path items. They could then use NLP to analyze project reports and identify which tasks are encountering the most delays or issues. By integrating these analyses, project managers can identify which tasks pose the highest risk of delaying the project and take measures to mitigate these risks.

Project managers can also leverage graph theory and NLP to identify opportunities for enhancement in project processes. For example, they might use graph theory to analyze the relationships among different teams or stakeholders involved in the project and identify areas of collaboration or conflict. They could then use NLP to analyze project reports and identify specific issues or challenges that are affecting these relationships. By addressing these issues, project managers can enhance team communication and collaboration, ultimately leading to improved project outcomes.

An additional method suggested is to employ graph theory and NLP to develop visualization tools that allow project managers to effortlessly monitor and analyze project data. For instance, project managers might utilize NLP to extract crucial information from project reports and present it in a visual format such as a network or graph. By using visualization tools, project managers can swiftly identify trends and patterns within the data and make informed decisions based on this information.

One potential application of using graph theory and NLP for agile project management is in the realm of software development. By constructing networks based on the relationships between different software components and using NLP to analyze project reports, project managers can gain insights into the dependencies and interactions between different components and identify areas for improvement. This can help ensure that the software development process is both efficient and effective.

Another potential application is in the realm of construction project management. By forming networks based on the relationships between various tasks and stakeholders involved in a construction project, project managers can gain insights into the dependencies and interactions between different components and pinpoint areas for improvement. This can help ensure that the construction project is completed within the stipulated time and budget.

In conclusion, the use of graph theory and NLP can offer valuable insights into the structure and dynamics of agile project management processes. Project managers can utilize these tools to identify patterns and trends within project data, pinpoint bottlenecks or vulnerabilities in the project, identify opportunities for improvement, and create visualization tools that enable real-time monitoring and analysis of project data. These methodologies have numerous potential applications in fields such as software development and construction project management, and can aid in ensuring that projects are completed on time and within budget. Fig. 5 consolidates all the steps of the machine learning workflow, presenting a detailed step-by-step overview.

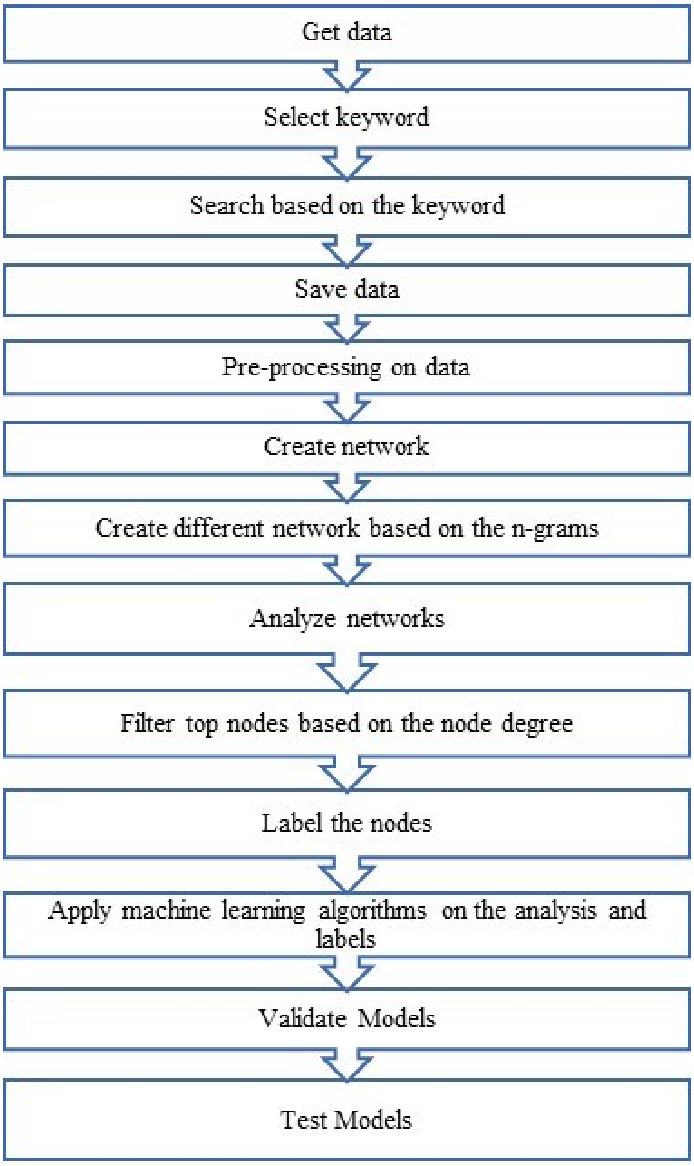


Figure 5. The machine learning workflow: From collection to analysis (own research)

4. RESULTS AND THEIR DISCUSSION

4.1. Data collection

Data collection is an indispensable component of research and decision-making processes across a wide array of domains, including academia, business, healthcare, and public policy. This paper explores the multifaceted importance of data collection by elaborating on its role in yielding valuable insights and its contribution to the comprehension of trends, patterns, and relationships. Additionally, the paper underscores the various sources and methods employed for data collection.

Data collection is a systematic procedure that involves the collection, measurement, and analysis of data from diverse sources. This process enables researchers and professionals to acquire empirical evidence that can inform their endeavors. Various sources of data collection are utilized, depending on the specific domain and research objectives. These sources include scientific journals, social media monitoring, online tracking, surveys, feedback mechanisms, and more.

One of the primary functions of data collection is the generation of valuable insights. By accumulating and analyzing data, researchers and decision-makers can gain a deeper understanding of the phenomena under investigation. These insights can unveil underlying trends, identify emerging patterns, and elucidate complex relationships, all of which are crucial for making informed decisions.

In the context of academic research, data collection assumes a central role in advancing knowledge and contributing to the scholarly discourse. As illustrated in the preceding chapter, the process of data collection in academia involves a diverse array of data types, encompassing both academic and non-academic data. This diversity of data sources allows researchers to explore various dimensions of a topic, thus enriching the research landscape.

Academic data collection involves the gathering of information related to academic disciplines, such as scientific experiments, survey responses, academic publications, and educational records. These data sources serve as the foundation for generating new knowledge, testing hypotheses, and validating theories. Researchers often rely on academic data to make significant contributions to their respective fields.

Non-academic data, on the other hand, extends beyond the traditional boundaries of academia. It encompasses data from sources such as social media, online reviews, customer feedback, and real-world observations. Non-academic data collection is particularly valuable for interdisciplinary research and for understanding the practical implications of academic findings in real-world scenarios. For academic data, tried to get the scientific articles from journals. Each paper consists of various metadata, like title, body, keywords, journal name, and etc. These data are the inputs for the next steps. All academic data stored in a single Json file, which each element in this file is the information about specific record.

As illustrated in the accompanying image, the dataset under consideration presents information pertaining to individual records. It is noteworthy that each record within this dataset offers a consistent set of data elements. While it is important to acknowledge that not every record within the dataset contains every available data element, a substantial portion of these records share commonalities, consistently encompassing the same set of information. To elucidate, some records may lack temporal data, yet the majority of records consistently feature both a title and a body section.

Within the context of academic data analysis, the title and body components of each record are subjected to meticulous filtering and are subsequently utilized as primary inputs for in-depth scrutiny and further analytical exploration. This choice of utilizing the title and body components stems from their intrinsic significance and recurring relevance within the dataset. These two components, namely the title and body, emerge as indispensable reservoirs of valuable insights, capable of furnishing users with a wealth of pertinent information. This intrinsic importance underscores the rationale for their selection as principal data sources for subsequent analysis and research endeavors.

The rationale behind this selection is grounded in the recognition that titles typically encapsulate succinct yet informative descriptors, serving as a brief but comprehensive representation of the record's content. On the other hand, the body section augments this by providing a more detailed exposition of the record's content, elucidating its nuances and intricacies. These dual components, when employed in tandem, synergistically enhance the analytical capabilities of researchers, facilitating the extraction of rich and meaningful insights from the dataset. Consequently, the combination of title and body data serves as a potent foundation for rigorous academic inquiry, offering researchers a robust starting point for their investigative endeavors and the potential to unlock a plethora of valuable information. Fig. 6 provides an example of the academic data information.


```
"LA": " English\n",
"DT": " Proceedings Paper\n",
"CT": " 10th Asia Pacific Structural Engineering and Construction Conference\n",
"CY": " NOV 13-15, 2018\n",
"CL": " Langkawi, MALAYSIA\n",
"SP": " Univ Teknologi Malaysia, Sch Civil Engn, Fac Engn, Construct Res Inst Malaysia\n",
"AB": " In Nigeria there is shortage of competent craftsmen due to project management skills (PMS) deficiency in their quality of
"C1": " [Inuwa, Ibrahim Ibrahim; Musa, Mohammed Mukhtar] Abubakar Tafawa Balewa Univ, Dept Quant Surveying, Bauchi, Nigeria.\n",
"RP": " Inuwa, II (corresponding author), Abubakar Tafawa Balewa Univ, Dept Quant Surveying, Bauchi, Nigeria.\n",
"EM": " iiinuwa@atbu.edu.ng\n",
"CR": " Adewale PO., 2014, INT J VOC TECH ED, V6, P36\n",
"NR": " 41\n",
"TC": " 0\n",
"Z9": " 0\n",
"U1": " 1\n",
"U2": " 1\n",
"PU": " IOP PUBLISHING LTD\n",
"PI": " BRISTOL\n",
"PA": " DIRAC HOUSE, TEMPLE BACK, BRISTOL BS1 6BE, ENGLAND\n",
"SN": " 1757-8981\n",
"J9": " IOP CONF SER-MAT SCI\n",
"PY": " 2019\n",
"VL": " 513\n",
"AR": " 012002\n",
"DI": " 10.1088/1757-899X/513/1/012002\n",
"PG": " 9\n",
"WC": " Construction & Building Technology; Engineering, Civil; Materials\n",
"WE": " Conference Proceedings Citation Index - Science (CPCI-S)\n",
"SC": " Construction & Building Technology; Engineering; Materials Science\n",
"GA": " BQ8UO\n",
"UT": " WOS:000621781200002\n",
"OA": " gold\n",
"DA": " 2022-12-03\n",
```


Figure 6. Academic data preview (own research)

The principal objective of this research endeavor is the application of advanced graph theory techniques and algorithms to the analysis of metadata, with particular emphasis on the title and body components. It is pertinent to acknowledge that the characteristics and attributes of the examined parameters may exhibit variations when comparing academic and non-academic datasets. A comprehensive discussion of these divergences is expounded upon within the respective sections dedicated to each dataset type.

Graph theory, as a mathematical framework, offers a robust and versatile analytical toolkit for the exploration of complex relationships and structures within data. By employing graph theory techniques, this study endeavors to harness the inherent power of this mathematical discipline to elucidate intricate patterns, associations, and dependencies that exist within the metadata of interest. Notably, the metadata elements of primary focus are the title and body, given their intrinsic significance in conveying information and context within textual data.

However, it is imperative to acknowledge that the nuances and characteristics of metadata can differ substantially between academic and non-academic contexts. These differences may manifest in terms of the nature of the information contained within the title and body, the language used, or the specific structural attributes unique to each dataset type. Therefore, this research recognizes the necessity of delineating and examining these distinctions comprehensively.

In the subsequent sections dedicated to academic and non-academic datasets, a meticulous exploration of the idiosyncrasies and variations within the metadata parameters is undertaken. This comparative analysis not only sheds light on the divergent attributes of the data but also contributes to a nuanced understanding of the specific challenges and opportunities associated with each dataset type.

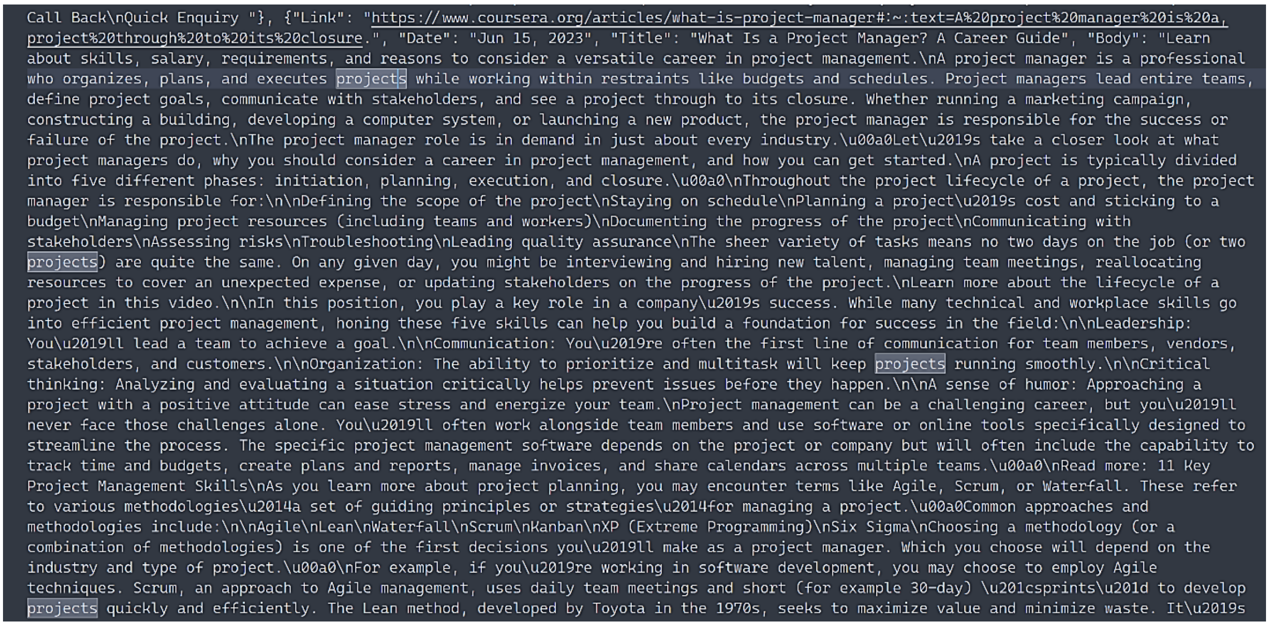
In the realm of data analysis, academic datasets exhibit a discernibly more structured and regular format in comparison to their non-academic counterparts. This characteristic owes its existence to a confluence of factors, primarily stemming from the meticulous curation and filtering mechanisms imposed by academic journals and their associated databases. This attribute of structuredness, as discussed in subsequent sections, bestows upon academic data a distinct advantage in terms of ease of processing, setting it apart from the relatively less structured landscape of non-academic data. While it is essential to acknowledge that the disparity between these two categories of data may not be seismic in nature, it undoubtedly contributes significantly to the enhanced manageability and utility of academic datasets within the domain of data analysis.

The structured nature of academic data, as previously alluded to, finds its origins in the rigorous editorial and review processes intrinsic to academic journals. These processes serve as a crucible for ensuring the quality, consistency, and organization of the information contained within academic publications. Additionally, academic databases, which often serve as repositories for scholarly works, further reinforce this structured paradigm by imposing standardized formatting and indexing practices. Consequently, academic data is imbued with a regularity and predictability that greatly simplifies its manipulation and analysis.

In contrast, non-academic data sources typically lack the stringent curation and filtering mechanisms characteristic of academic data. This divergence can lead to a higher degree of heterogeneity, unpredictability, and noise within non-academic datasets. The absence of a standardized framework for organizing and presenting information can present formidable challenges when attempting to process and extract meaningful insights from such data.

It is important to underscore that while the disparity in structure between academic and non-academic data may not always result in a profound divergence in terms of analytical outcomes, it undeniably expedites the workflow associated with academic data analysis. This increased ease of use can translate into efficiency gains, facilitating more expedient research and knowledge generation processes within academic contexts.

In the Fig. 7 a record of non-academic data has been shown. As obvious from the picture each result consists of four elements (i.e., Link, Date, Title, and Body). Also, it can be mentioned these data is not as clean as academic data.



```
Call Back\nQuick Enquiry "}, {"Link": "https://www.coursera.org/articles/what-is-project-manager#:~:text=A%20project%20manager%20is%20a,
project%20through%20to%20its%20closure.", "Date": "Jun 15, 2023", "Title": "What Is a Project Manager? A Career Guide", "Body": "Learn
about skills, salary, requirements, and reasons to consider a versatile career in project management.\nA project manager is a professional
who organizes, plans, and executes projects while working within restraints like budgets and schedules. Project managers lead entire teams,
define project goals, communicate with stakeholders, and see a project through to its closure. Whether running a marketing campaign,
constructing a building, developing a computer system, or launching a new product, the project manager is responsible for the success or
failure of the project.\nThe project manager role is in demand in just about every industry.\u00a0Let\u2019s take a closer look at what
project managers do, why you should consider a career in project management, and how you can get started.\nA project is typically divided
into five different phases: initiation, planning, execution, and closure.\u00a0\nThroughout the project lifecycle of a project, the project
manager is responsible for:\n\nDefining the scope of the project\nStaying on schedule\nPlanning a project\u2019s cost and sticking to a
budget\nManaging project resources (including teams and workers)\nDocumenting the progress of the project\nCommunicating with
stakeholders\nAssessing risks\nTroubleshooting\nLeading quality assurance\nThe sheer variety of tasks means no two days on the job (or two
projects) are quite the same. On any given day, you might be interviewing and hiring new talent, managing team meetings, reallocating
resources to cover an unexpected expense, or updating stakeholders on the progress of the project.\nLearn more about the lifecycle of a
project in this video.\n\nIn this position, you play a key role in a company\u2019s success. While many technical and workplace skills go
into efficient project management, honing these five skills can help you build a foundation for success in the field:\n\nLeadership:
You\u2019ll lead a team to achieve a goal.\n\nCommunication: You\u2019re often the first line of communication for team members, vendors,
stakeholders, and customers.\n\nOrganization: The ability to prioritize and multitask will keep projects running smoothly.\n\nCritical
thinking: Analyzing and evaluating a situation critically helps prevent issues before they happen.\n\nA sense of humor: Approaching a
project with a positive attitude can ease stress and energize your team.\nProject management can be a challenging career, but you\u2019ll
never face those challenges alone. You\u2019ll often work alongside team members and use software or online tools specifically designed to
streamline the process. The specific project management software depends on the project or company but will often include the capability to
track time and budgets, create plans and reports, manage invoices, and share calendars across multiple teams.\u00a0\nRead more: 11 Key
Project Management Skills\nAs you learn more about project planning, you may encounter terms like Agile, Scrum, or Waterfall. These refer
to various methodologies\u2014a set of guiding principles or strategies\u2014for managing a project.\u00a0Common approaches and
methodologies include:\n\nAgile\nLean\nWaterfall\nScrum\nKanban\nXP (Extreme Programming)\nSix Sigma\nChoosing a methodology (or a
combination of methodologies) is one of the first decisions you\u2019ll make as a project manager. Which you choose will depend on the
industry and type of project.\u00a0\nFor example, if you\u2019re working in software development, you may choose to employ Agile
techniques. Scrum, an approach to Agile management, uses daily team meetings and short (for example 30-day) \u201csprints\u201d to develop
projects quickly and efficiently. The Lean method, developed by Toyota in the 1970s, seeks to maximize value and minimize waste. It\u2019s
```


Figure 7. Non-Academic data preview (own research)

The methodological approach for this non-academic data gathering operation involved a series of meticulously designed steps, each contributing to the overall objective of accumulating a meaningful dataset on project management.

The initial step revolved around the careful selection of a keyword that would serve as the foundation of the search process. 'Project management' was chosen due to its broad relevance and significance in various professional domains. This keyword was expected to yield diverse sources of information, encompassing the latest trends, practices, and discussions within the field.

The Google search engine was identified as the optimal tool for data collection. Google's unparalleled indexing capabilities and user-friendly interface made it the ideal choice for sourcing web-based data. Furthermore, its advanced search features allowed us to fine-tune results, which proved pivotal in achieving the research's temporal objectives.

To ensure the relevance and timeliness of the acquired data, I implemented a specific date range for the search results. This date range was set to span from January 1, 2023, to July 31, 2023, reflecting the desire to capture the most recent developments and discussions in project management. By limiting the scope to this seven-month period, I sought to balance the need for current data with the requirement for a substantial dataset for analysis.

Each webpage discovered during the search process was recognized as a repository of valuable information. To harness this data effectively, I systematically extracted key elements from each webpage, including the title, body text, and publication date. These elements were considered fundamental for constructing a dataset that would support comprehensive analysis and insights.

Despite the methodical approach, several challenges emerged during the data gathering process, each requiring careful consideration and resolution:

An initial challenge lay in the absence of publication dates on certain webpages. Without this temporal information, it was difficult to assess the relevance of the content. To address this issue, I relied on alternative strategies, such as searching for publication date metadata within the webpage's source code or cross-referencing information with other credible sources.

This non-academic data gathering venture resulted in the acquisition of a substantial dataset encompassing a diverse array of web-based resources related to project management. The data, comprising titles, body text, and, where available, publication dates, holds the potential to serve as a valuable resource for future analysis, research, and insights into the evolving landscape of project management practices. Furthermore, the challenges encountered during this endeavor have provided valuable lessons and insights into the complexities of web-based data collection, which can inform future non-academic research efforts in the digital era.

Datasets from historical records are crucial in guiding research, influencing decision-making processes, and comprehending temporal trends. This academic text delves into the complexities encountered in the pursuit of constructing a historical dataset for non-academic purposes. The objective was to create 10 to 12 distinct time periods, each containing at least 1000 records, focusing on the field of project management. This endeavor required data collection from web sources and involved a multi-faceted approach, beset with challenges and innovative solutions.

One of the initial and persistent challenges was the presence of webpage restrictions designed to deter data scraping. Some websites employed mechanisms such as CAPTCHA challenges and robots.txt files, making it difficult to extract the desired data elements, including titles and body text. These restrictions necessitated the exploration of workarounds and alternative data sources to compensate for the missing data.

The scale of search results generated by the Google search engine posed a significant challenge. Typically, the search results included around 30 pages, with each page containing approximately 10 webpages. This abundance of data required a structured approach to ensure comprehensive coverage. Consequently, the decision was made to divide the search results into three stages, with each month further divided into three date ranges. While this strategy aimed to maximize data collection, it introduced its own set of challenges.

Despite meticulous planning and dividing the search results into date ranges, an unexpected challenge emerged. Some months did not yield the desired volume of data necessary to meet the 1000-record threshold. This disparity raised questions about the completeness and representativeness of the dataset for each month.

To address webpage restrictions, the research team employed a multi-pronged approach. First, CAPTCHA-solving solutions were explored to bypass challenges encountered during data scraping. Second, alternative data sources were identified to compensate for inaccessible webpages. These included publicly available datasets and archived content from websites with more permissive data policies. The integration of these supplementary sources helped mitigate the impact of webpage restrictions.

To cope with the overwhelming volume of search results, a structured three-stage data collection process was adopted. Each month was divided into three distinct date ranges, allowing for more granular searches. This approach enabled the research team to obtain a more comprehensive dataset while managing the scale of information. Despite this, the challenge of ensuring an equitable distribution of data across all months persisted.

In response to the issue of insufficient data for some months, a strategic decision was made to aggregate records from each consecutive three-month period into a single period. This consolidation yielded ten distinct periods, each spanning three months. By grouping data in this manner, the research team aimed to achieve the desired minimum of 1000 records per period. While this approach helped address the challenge of data scarcity in individual months, it required careful consideration to ensure meaningful analysis could still be conducted.

As mentioned, the non-academic data collected from web. Therefore, I used Python to automate this task, especially, used requests and selenium. The Requests library, a versatile Python library for handling HTTP requests, played a fundamental role in the initial data collection phase. It was employed to initiate HTTP GET requests to the Google search engine, where the query term 'project management' was used to fetch search results. Requests' efficiency in sending HTTP requests and processing server responses enabled the rapid retrieval of links to webpages. Furthermore, Requests was pivotal in extracting the publication dates associated with each webpage. This operation involved the parsing of the HTML content of the search results, locating and extracting date information embedded within the source code. Requests' simplicity and speed made it an ideal tool for this initial data retrieval task.

Selenium, renowned for its web automation capabilities, assumed a crucial role in the subsequent phases of data collection. Selenium was employed for its unique ability to interact with dynamic web content, execute user-like interactions with webpages, and extract structured data. In this project, Selenium was used to directly access and scrape webpage contents, focusing on the extraction of titles and body text. Its strengths lay in its adaptability to a variety of webpage layouts and its capacity to navigate webpages with complex structures and interactive elements. Selenium's dynamic interaction capabilities, including finding HTML elements, simulating user actions, and capturing data, proved essential for the comprehensive collection of textual data from webpages.

During the initial phase, Requests was employed to initiate HTTP GET requests to the Google search engine, where the search query 'project management' was executed. The search results, comprised of a diverse array of webpages, were retrieved as HTML content. Concurrently, Requests was used to extract the publication dates associated with each webpage. This process involved parsing the HTML content of the search results, employing techniques such as Regular Expressions or parsing libraries, to locate and extract date information embedded within the HTML source code. This step was critical for establishing temporal context within the dataset, ensuring that the collected data was accurately time-stamped.

Following the acquisition of links and dates, Selenium took center stage. Selenium was employed to dynamically interact with webpages and extract the desired content, including titles and body text. Selenium's flexible nature allowed it to navigate through webpages, locate HTML elements, and simulate user actions, enabling the retrieval of textual data. Its adaptability to diverse webpage structures and its ability to handle JavaScript-driven interactions made Selenium an indispensable tool for the comprehensive collection of webpage data. The process involved initiating Selenium WebDriver sessions, opening webpages, navigating through webpage structures, and extracting data elements based on CSS selectors or XPath expressions.

4.2. NLP implementation

Once the data extraction phase concluded, the collected data elements—publication dates, titles, and body text—were systematically organized and stored in a unified JSON (JavaScript Object Notation) file. JSON's lightweight and human-readable format facilitated the structured storage of data. Each webpage's data was typically represented as a JSON object, allowing for easy organization and retrieval. This consolidation step aimed to provide a unified repository for the acquired information, simplifying subsequent data analysis processes.

Subsequent to the acquisition of both academic and non-academic datasets, my research embarked on a crucial phase known as data preprocessing. This phase is fundamentally essential in the domain of data science, particularly in the context of Natural Language Processing (NLP). Its primary objective is to transform raw and often unstructured data into a structured and refined format that is amenable to rigorous analysis and modeling.

Data preprocessing is a multi-faceted process that encompasses several intricate steps, each aimed at addressing specific challenges inherent in the data. One of the foundational steps in this process, which warrants a deeper exploration, is tokenization.

Tokenization is a fundamental and indispensable technique in NLP. It serves as the initial step in the transformation of unstructured text data into a format that can be comprehensively analyzed and utilized in various NLP applications. The essence of tokenization lies in the segmentation of lengthy and continuous text strings into smaller, discrete units known as tokens.

Tokens can take various forms, including words, symbols, numerals, punctuation marks, and any other linguistic elements present in the text. By segmenting the text into tokens, I effectively dissect it into its constituent parts, creating the building blocks that facilitate my understanding of its linguistic context.

In my specific research context, tokenization was carried out using the "blank space" as the delimiter. This strategy involved dividing the text string into tokens wherever a space character was encountered. This choice of delimiter was made after careful consideration of the nature of my data and the specific objectives of my study.

The selection of an appropriate tokenization strategy is not a trivial matter and can have a significant impact on the subsequent analysis. Different delimiters can be chosen based on the characteristics of the data and the specific research questions being addressed. For instance, in some cases, more advanced tokenization techniques may be employed, such as sentence or word-level tokenization, which can be particularly valuable for certain NLP tasks like sentiment analysis or machine translation.

Here, I delve into two essential techniques that play a pivotal role in this phase—stemming and lemmatization. These techniques are instrumental in addressing the inherent intricacies and variations in word forms that pervade the English language, ultimately endeavoring to distill words to their purest, elemental forms for the purpose of enabling more profound and effective text analysis.

To appreciate the significance of these techniques, let us consider a practical scenario. Imagine a user engaging in a product search on a prominent e-commerce platform like Amazon. The user enters the query "shirts" into the search bar, anticipating not only exact matches but also results that encompass related word forms, such as "shirt." The English language, with its rich tapestry of expressions, often presents words in diverse forms dictated by tense, context, or grammatical usage. For instance, the verbs "go," "going," and "went" all share a common etymological root, yet their outward appearances diverge significantly based on subtle contextual nuances. This multiplicity of word forms poses a formidable challenge when it comes to achieving comprehensive information retrieval.

Stemming presents one approach to surmounting this challenge. It relies on a heuristic process aimed at reducing words to their foundational forms by judiciously pruning word endings. However, it is imperative to acknowledge that stemming operates on rule-based heuristics, scanning for and excising common suffixes associated with word variations—examples include "-ed" or "-ing" (e.g., "asked" and "asking"). While stemming may offer computational efficiency, its efficacy is inherently constrained by its propensity to yield word forms that may or may not constitute meaningful base words, especially when confronted with the idiosyncrasies of English.

In stark contrast, lemmatization emerges as a more sophisticated and linguistically informed technique. It aspires to confront word variation with surgical precision by harnessing an extensive vocabulary and engaging in morphological analysis of words. Through this methodical process, lemmatization consistently begets the fundamental or dictionary form of a word—a form known as the lemma. Notably, lemmatization shines in its versatility, reliably reducing even the most irregular words, such as "mice" and "ran," to their respective lemmas. This showcases its linguistic acumen and adaptability in the face of the English language's manifold idiosyncrasies.

In the context of my research endeavors, I made a conscious decision to harness the power of lemmatization as the post-tokenization technique of choice for my dataset. This decision was underpinned by my recognition that the English language, replete with its inherent ambiguities, benefits profoundly from the nuanced and comprehensive approach afforded by lemmatization. By implementing lemmatization, my aim was to cultivate a dataset characterized by refined and contextually precise word representations, thereby raising the quality and effectiveness of subsequent NLP analyses.

One such critical step that follows stemming or lemmatization is the removal of stop words. This procedure plays a pivotal role in fine-tuning textual data for subsequent analysis, as it seeks to eliminate words that serve as linguistic "fillers" and bear minimal intrinsic meaning. These linguistic fillers predominantly encompass conjunctions such as "because," "and," and "since," as well as prepositions like "under," "above," "in," and "at." While these words constitute a substantial portion of human language, they often lack substantive relevance when developing NLP models.

It is worth noting that the removal of stop words is not an absolute necessity in every NLP modeling scenario; its applicability hinges upon the specific task at hand. For instance, when conducting text classification, where the objective is to categorize text into different genres, filter out spam, or generate automated tags, eliminating stop words from the text proves advantageous. This strategic elimination enables the model to focus its attention on words that convey the essence and meaning of the text within the dataset. Conversely, in tasks such as text summarization and machine translation, the removal of stop words may not be warranted, as these words often contribute to sentence structure and coherence.

Multiple methods and libraries exist to facilitate the removal of stop words, including renowned NLP libraries such as Gensim, SpaCy, and NLTK (Natural Language Toolkit). In my research endeavor, I leveraged the capabilities of the NLTK library to effectuate the removal of stop words from my dataset. NLTK's comprehensive stop words corpus allows for the selective removal of these non-essential linguistic components, streamlining the textual data for more nuanced and contextually relevant analyses.

Subsequently, following the data cleaning process, a meticulously organized dataset emerges, setting the stage for the crucial task of word frequency calculation. Various techniques are available for achieving this objective, with two prominent methodologies frequently employed within the realm of natural language processing: TF-IDF Vectorization and Count Vectorization. These techniques serve as means to transform textual data into numerical representations for further analysis. However, it is imperative to acknowledge a fundamental distinction between these two approaches.

The TF-IDF (Term Frequency-Inverse Document Frequency) Vectorizer is a pivotal tool in the field of text analysis, renowned for its capacity to offer a nuanced understanding of textual data. At its core, TF-IDF Vectorizer operates on the fundamental premise that not all words within a document are equally informative or significant. This technique aims to address this issue through a multifaceted approach.

Term Frequency (TF): TF measures the frequency with which a word appears within a specific document. Words that occur more frequently within a document are assigned higher TF values, indicating their prominence within that particular textual context.

Inverse Document Frequency (IDF): IDF, on the other hand, evaluates the uniqueness of a word across the entire corpus of documents. Words that are rare across the corpus receive higher IDF values, signifying their potential importance in distinguishing one document from another.

The amalgamation of TF and IDF produces a composite score for each word within a document. This score encapsulates the word's importance or significance within that specific document relative to the entirety of the corpus. In essence, TF-IDF provides a nuanced metric that quantifies how crucial a word is to a given document or sentence.

The significance of TF-IDF lies in its ability to differentiate between words that are merely frequent and those that hold substantial meaning within the broader context of the dataset. This nuanced approach is particularly advantageous in various natural language processing tasks, including but not limited to document classification, information retrieval, and text summarization. TF-IDF excels in capturing the semantic relevance of words and is an invaluable tool for discerning the salience of terms within a text.

Count Vectorizer, in contrast to the complexity of TF-IDF Vectorizer, offers a straightforward approach to text analysis. It serves as a text-to-numerical conversion tool, transforming a collection of text documents into a matrix that quantifies the raw frequency of tokens (words or terms). Unlike TF-IDF, Count Vectorizer does not consider the broader context of the entire corpus when assigning values to words. Instead, it merely tallies how many times each word appears within a specific document.

The resultant matrix generated by Count Vectorizer provides a direct and interpretable representation of textual data. In this matrix, each row corresponds to a document, and each column corresponds to a unique word or term. The values contained within the matrix denote the raw frequency of each word within its respective document.

A fundamental distinction between TF-IDF Vectorizer and Count Vectorizer lies in their computational efficiency and complexity. TF-IDF Vectorizer, while providing a rich perspective on term importance within a document, can be computationally demanding. This becomes evident when dealing with complex linguistic constructs, such as calculating the frequency of word combinations (e.g., doubles and triples). To do so, the algorithm needs to consider both term frequency and inverse document frequency for every word in every document, which can slow down processing, especially for large datasets.

Count Vectorizer, on the other hand, is notably simpler in its approach. It excels in terms of computational speed and efficiency, particularly when dealing with tasks involving the analysis of multiple word combinations. This advantage becomes especially apparent when working with extensive datasets or when employing the technique in machine learning algorithms, where computational efficiency can be a critical factor.

In summary, the choice between TF-IDF Vectorizer and Count Vectorizer hinges on specific text analysis objectives and dataset characteristics. While TF-IDF offers context-awareness and nuanced term importance metrics, Count Vectorizer prioritizes computational efficiency and simplicity. Understanding these differences empowers practitioners to make informed decisions about which technique best aligns with their analytical goals in the dynamic field of text analysis and natural language processing.

In this study for generating the co-occurrence matrix I used the Count Vectorizer in the scikit-learn python module. It has been tried to explain the procedure for creating a co-occurrence matrix using Count Vectorizer is as follows:

- **Import the required libraries:** Begin by importing the necessary libraries in your programming environment. You'll need the sklearn.feature_extraction.text module from scikit-learn, which provides the Count Vectorizer class for text feature extraction, and the numpy library for working with matrices.
- **Prepare your corpus:** A corpus is a collection of documents or sentences that you want to analyze. It could be your own dataset or sample text data. Make sure your corpus is in a format that can be processed by CountVectorizer.
- **Create an instance of CountVectorizer:** Initialize an instance of the CountVectorizer class. This class is responsible for converting text documents into a matrix of token counts. By default, it tokenizes the input text on white spaces and converts everything to lowercase.
- **Fit the CountVectorizer on corpus:** Use the fit_transform () method of the CountVectorizer instance to fit the model on your corpus. This step analyzes the text data and builds the vocabulary of unique tokens present in your corpus. It also assigns a numerical index to each token. In this step I can adjust the size of the ngram.
- **Generate the document-term frequency matrix:** Once the CountVectorizer is fitted on your corpus, use the transform () method to generate the document-term frequency matrix. This matrix represents the number of times each token appears in each document of the corpus.

- **Compute the co-occurrence matrix:** To create a co-occurrence matrix, you need to perform a mathematical operation on the document-term frequency matrix. The simplest approach is to multiply the matrix by its transpose, which results in a square matrix where each element reflects the co-occurrence count of two tokens within the same context (document).
- **Interpret the co-occurrence matrix:** The resulting co-occurrence matrix will have rows and columns corresponding to the unique tokens in your corpus. The value at position (i, j) in the matrix represents the number of times token i and token j co-occur within the same document.

Unigrams, bigrams, and trigrams are different approaches to representing and analyzing text data, each with its own advantages and use cases. The decision on the superior method is contingent upon the particular task at hand and the attributes of the data set.

Unigrams, or individual words or tokens, are a fundamental way of representing text data in natural language processing. They offer several advantages due to their simplicity and versatility. Unigrams are easy to work with, computationally efficient, and highly interpretable. Each word is treated as a separate entity, making it straightforward to analyze and understand text-based information.

One of the key advantages of unigrams is their efficiency in processing and analysis, especially in large datasets. Their simplicity allows for faster training and inference times compared to more complex text representations. Unigrams are also highly interpretable because they directly represent individual words, making it easy to extract insights and draw conclusions from text data. This interpretability is especially valuable in tasks like sentiment analysis, topic modeling, and text classification, where understanding the contribution of individual words is crucial.

Unigrams are versatile and can be used for a wide range of NLP tasks. They serve as a baseline representation for text data, providing a simple yet reasonable starting point for experimenting with more advanced techniques and models. They can be preprocessed easily by applying techniques like stemming, lemmatization, or stop word removal, which helps improve data quality by reducing noise. Additionally, unigrams can be used as features in machine learning models, allowing for the creation of informative features for various text-based applications.

However, it's important to note that unigrams have limitations, particularly in capturing semantic meaning and context. They don't consider word order or relationships between words, which can be critical for certain NLP tasks. For tasks requiring a deeper understanding of language semantics and context, more advanced text representations like bigrams, trigrams, word embeddings, or deep learning-based models (e.g., Transformers) may be necessary. The choice of representation ultimately depends on the specific task, dataset characteristics, and the level of linguistic complexity required for the analysis. Unigrams remain a valuable and efficient option for many text analysis scenarios, offering a straightforward and interpretable approach to handling textual data while being aware of their limitations.

Bigrams, which are pairs of adjacent words, are essential components in natural language processing and text analysis. They play a crucial role in capturing contextual information within text data. This context is vital because the meaning of many words in a sentence depends on the words that come before or after them. For instance, the phrase "hot dog" has a completely different meaning than "hot weather" or "dog food." Bigrams help us understand these subtle nuances in language by considering the specific word pairs in a text. This contextual information is invaluable in various natural language processing tasks.

One significant application of bigrams is phrase identification. Bigrams enable the identification of common phrases and expressions within text. This is particularly useful in sentiment analysis, where the presence of certain bigrams can significantly affect sentiment classification. For example, identifying the bigram "not good" can be pivotal in accurately classifying a sentence as having a negative sentiment, whereas individual words like "not" and "good" might not convey the same sentiment when considered separately. Bigrams provide a more granular view of language, allowing for more precise sentiment analysis.

In text classification and information retrieval tasks, bigrams serve as discriminative features. They help distinguish between documents or texts more effectively than individual words. By considering pairs of adjacent words, bigrams provide richer information about the content and semantics of a text. This discriminative power enhances the accuracy of classification algorithms and information retrieval systems. This approach enables a more detailed comprehension of text, facilitating the distinction between documents that may share similar individual words but possess different meanings in context.

Furthermore, bigrams can help reduce noise in certain languages where single words can be highly ambiguous. In such languages, the meaning of a word may depend on its context, and a single word can have multiple senses. Bigrams come to the rescue by considering the words that surround a given word, thus disambiguating its meaning. This feature is particularly beneficial in languages with complex word meanings and multiple word senses, as it helps improve the precision of various natural language processing tasks.

Trigrams represent a notable advancement in NLP due to their unique ability to provide a deeper and more intricate context when analyzing text data. Unlike unigrams, which focus on individual words, and bigrams, which consider pairs of consecutive words, trigrams take into account three consecutive words in a text sequence. This extended context allows NLP models to gain a more profound understanding of the intricate dependencies and relationships that exist among words within sentences and documents.

In various NLP applications, having precise context is of paramount importance. Trigrams shine in tasks where this heightened level of context is crucial. Consider machine translation, for example, where accurately capturing the context of three consecutive words can significantly improve the choice of translation for a given phrase or sentence. This leads to more accurate and contextually relevant translations, benefiting automated translation services and their users alike. Similarly, in speech recognition, understanding the context of the preceding and following words can greatly enhance the accuracy of transcriptions, ensuring that spoken words are transcribed more faithfully.

One of the standout strengths of trigrams lies in their capability to capture not just common phrases but also complex syntactic structures and intricate language patterns. They excel at identifying and dissecting grammatical constructs and revealing subtle relationships between words. For instance, when confronted with a sentence like "The cat chased the mouse," a trigram model can discern the precise syntactic structure: "the cat" as the subject, "chased" as the verb, and "the mouse" as the object. This level of granularity in analysis surpasses what unigrams or bigrams can offer, making trigrams particularly valuable in applications like syntactic parsing and semantic role labeling.

Trigrams also play a pivotal role in named entity recognition (NER) tasks. In NER, the goal is to identify and classify entities within text, such as names of people, places, organizations, and more. Trigrams excel in this domain because they are adept at recognizing multi-word entities. For instance, identifying "New York City" as a single entity can be challenging for unigrams, but trigrams excel at capturing such multi-word entities accurately. This is crucial in applications like information extraction, where precise identification of entities is vital for structuring and organizing data.

In the realm of language modeling, trigrams offer substantial advantages. When it comes to predicting the next word in a sequence, having a richer and more extensive context, as provided by trigrams, results in more accurate and contextually relevant text generation. This is especially crucial for applications like chatbots, text completion, and text generation tasks, where generating coherent and contextually appropriate responses is essential for improving user interactions and experiences.

Nonetheless, it is crucial to recognize that trigrams present unique challenges. The main compromise is a surge in dimensionality, as the size of the vocabulary and the potential three-word combinations increase substantially. This can result in data sparsity problems, where certain trigrams may appear rarely or not at all in a specific dataset. Data sparsity can affect the computational efficiency and overall performance of models.

Hence, when choosing between unigrams, bigrams, or trigrams, it is vital to take into account the specific needs of the NLP task and the attributes of the dataset being used. A recommended approach is to experiment with various n-gram models and assess their performance. This iterative process helps to find a balance between the depth of context provided by trigrams and the available computational resources, ensuring that the NLP system is fine-tuned for the particular application.

In light of the merits and limitations associated with various n-gram configurations, I systematically incorporated all available n-grams into my dataset analysis. Within the scope of my research, the utilization of both bigrams and trigrams proved to be particularly instrumental in unearthing a wealth of invaluable insights and intricate patterns. In both datasets, I undertook the generation of n-grams, including unigrams, bigrams, and trigrams, specifically targeting the textual components of "Title" and "Body."

In my analytical approach, I found it essential to differentiate between these two textual components based on their inherent characteristics and the nature of the data they encompass. In the "Body" section, characterized by a substantial volume of text and data, I opted for the utilization of the count vectorizer. This choice was motivated by the imperative of simplicity and efficiency, which count vectorization offers in handling extensive textual data. Given the sheer magnitude of text in this section, this approach excelled in swiftly processing and extracting relevant information, surpassing the performance of the Tf-IDF (Term Frequency-Inverse Document Frequency) method. The computational efficiency of the count vectorizer proved invaluable, allowing us to manage and analyze large volumes of text data effectively.

Conversely, in the "Title" section, where the quantity of textual content is considerably lower and computational resources are not a limiting factor, the choice between count vectorization and Tf-IDF did not yield a substantial difference in performance. The relative simplicity of the "Title" data, coupled with an absence of computational constraints, meant that both methods were equally capable of handling the task at hand. Therefore, in this context, the choice between count vectorization and Tf-IDF did not yield a significant impact on my results.

My approach, which systematically incorporated these n-gram features, facilitated a thorough examination of the data landscape. It allowed us to unveil hidden correlations, thematic nuances, and semantic intricacies that might have eluded conventional analyses. As I delve into the subsequent sections, I will elaborate on the specific advantages and insights gained through this strategic utilization of bigrams and trigrams, shedding light on the multifaceted contributions they made to my study.

The outcome generated by the count vectorizer does not yield a co-occurrence matrix directly. In the realm of natural language processing and text analysis, a co-occurrence matrix is a specialized square matrix wherein each row and column correspond to a distinct linguistic element, be it a word or an n-gram, and the cell residing at the intersection of a row and column contains a numerical value denoting the magnitude or frequency of co-occurrence between the respective elements. Consequently, to derive the co-occurrence matrix from the count vectorizer's output, certain transformations must be applied.

Initially, the count vectorizer matrix undergoes a transposition operation. This transposed matrix subsequently assumes a structure wherein its rows now represent the distinct n-grams encountered in the corpus, while the columns correspond to individual documents or segments of text. The purpose of this transposition is to facilitate the alignment of n-grams with documents, as it ensures that the rows are associated with unique linguistic elements.

Subsequently, a further transformation is performed by multiplying the transposed matrix by itself. This mathematical operation serves as the cornerstone for obtaining the co-occurrence matrix. In practical terms, this multiplication operation calculates the co-occurrence relationships between every n-gram and all other n-grams present in the corpus. This comprehensive assessment of co-occurrence patterns is pivotal in various natural language processing tasks, including semantic analysis, document clustering, and sentiment analysis, as it elucidates the underlying semantic relationships and contextual associations between linguistic elements within the text data. This rigorous methodological approach ensures a more profound understanding of the intricate interplay of elements in textual data analysis, thereby enhancing the capabilities of computational linguistic models and advancing the field's research endeavors.

4.3. Network creation

Subsequent to the rigorous process of n-gram generation, a critical phase of my research unfolded as I embarked on the construction of network models built around these n-grams. This undertaking was executed with meticulous precision through the utilization of the Python programming language's network module, a versatile toolset that empowers the creation of network structures to encapsulate intricate linguistic relationships.

To create the network representation, I can use either an adjacency matrix or an edge list:

Adjacency Matrix: An adjacency matrix is a square matrix where the rows and columns correspond to the n-grams, and the cell values represent the co-occurrence weights. Each cell (i, j) in the adjacency matrix contains the weight or strength of the co-occurrence relationship between the corresponding n-grams (nodes).

Edge List: An edge list is a tabular representation where each row represents an edge in the network. It typically includes information about the source node, target node, and the weight of the edge. In this case, the source and target nodes will be n-grams, and the weight will be the co-occurrence value from the co-occurrence matrix.

Once the network is created using either the adjacency matrix or the edge list, it can be further analyzed, visualized, or utilized for various tasks such as word association analysis, community detection, or recommendation systems.

My methodological approach involved the creation of a dedicated Graph for each distinct n-gram present within my comprehensive dataset. These n-grams, serving as the foundational building blocks, held the capacity to encompass a wide spectrum of linguistic granularity, ranging from the fundamental unigrams (single words) to the more complex bigrams and trigrams. In essence, each node within these network structures represented a singular word or a specific n-gram, thereby facilitating the encapsulation of a rich array of linguistic information.

The edges connecting these nodes were paramount in defining the core essence of the network. These edges represented the relationships, or more precisely, the cooccurrence patterns between the nodes. The concept of cooccurrence, here, served as the linchpin of my network's construction. It enabled us to encode the frequency and proximity with which specific words or n-grams occurred together within my corpus. Consequently, these edges became conduits of meaning, vividly depicting the interwoven fabric of language, where nodes connected when they shared contextual significance.

Having established these intricate network structures, the subsequent step in my analytical pipeline involved the visual representation of these networks. Visualization served as a potent means to unlock essential insights into the inherent structural characteristics of these complex graphs. Through visualization, I could explore the intricacies of these linguistic networks in a manner that transcended raw numerical data. I could perceive clusters of nodes, identify central hubs, and observe patterns of connectivity. Moreover, visualization bestowed upon us a critical means of simplifying and communicating these complex relationships to both experts and non-experts alike.

Once a graph is created, it becomes possible to generate a succinct summary of the network for each dataset. This summary provides valuable insights into the graph's characteristics. For instance, if I examine the accompanying image, I can observe a comprehensive overview of the graphs presented. The summary encompasses crucial details such as the number of nodes and edges within the network, allowing us to understand its size and complexity. Additionally, it includes information about the density of the graph, which indicates how interconnected the nodes are. Furthermore, the average degree of the entire network is provided, offering insights into the average number of connections or neighbors each node possesses. By analyzing these key metrics, I gain a deeper understanding of the underlying structure and properties of the network at hand.

Similar to Fig. 8, we can provide a summary of the main information regarding the networks.

Academic Keywords	Academic Titles	Academic Abstract
Number of nodes: 214 Number of edges: 1271 Density: 0.05576762757228731 Average degree: 11.878504672897197	Number of nodes: 3276 Number of edges: 24039 Density: 0.004481167687274558 Average degree: 14.675824175824175	Number of nodes: 11183 Number of edges: 2537692 Density: 0.040587387085435606 Average degree: 453.848162389341
Number of nodes: 441 Number of edges: 1526 Density: 0.015728715728715727 Average degree: 6.920634920634921	Number of nodes: 10333 Number of edges: 17522 Density: 0.0003282485715042548 Average degree: 3.3914642407819606	Number of nodes: 155851 Number of edges: 10453684 Density: 0.0008607617576182224 Average degree: 134.14971992479997

Figure 8. Details of some generated networks for academic data (own research)

The visualization of complex networks, characterized by a substantial volume of data and a multitude of interconnected nodes, often poses a significant challenge in terms of clarity and comprehensibility. In such scenarios, it becomes imperative to employ strategic techniques to enhance the visualization process.

To address this challenge, one common approach is to apply filtering techniques to selectively include only the most relevant nodes and edges in the visualization. This filtering process aims to highlight the core elements of the graph while suppressing or omitting less significant components. In my study, I specifically focused on filtering nodes based on their topological characteristics, namely their degree within the graph.

The degree of a node in a graph refers to the number of edges connected to that node. Nodes with high degrees typically represent hubs or central elements within the network, while nodes with low degrees are often peripheral or less influential. Leveraging this information, I applied a filtering criterion to select the top nodes based on their degrees. In my case, I opted to visualize only the top 100 nodes, thereby reducing the complexity of the graph and enhancing the clarity of the resulting visualization.

This selective approach to node inclusion is underpinned by the assumption that nodes with higher degrees are more critical to understanding the overall network structure and dynamics. By focusing on these top nodes, I aimed to distill the essential information from the graph while relegating less influential nodes to the background. This strategy not only improves the visual aesthetics of the graph but also simplifies the task of interpreting and extracting insights from the visualization.

Moreover, the process of filtering out less important nodes based on degree also aligns with the underlying characteristics of the dataset. Nodes with lower degrees may be less prevalent or less interconnected, making them less informative for the specific analysis at hand. Consequently, by excluding these nodes, I aimed to create a more informative and actionable visualization that emphasizes the network's core components.

4.4. Network visualization

Academic data Keywords Unigrams:

15 top nodes based on degree:

[('science', 127), ('engineering', 77), ('multidisciplinary', 64), ('computer', 62), ('business', 8), ('research', 45), ('interdisciplinary', 41), ('economics', 40), ('system', 39), ('theory', 39), ('environmental', 38), ('method', 38), ('electrical', 37), ('study', 37), ('applied', 36)].

The detailed visualization of the network is presented in Fig. 9.

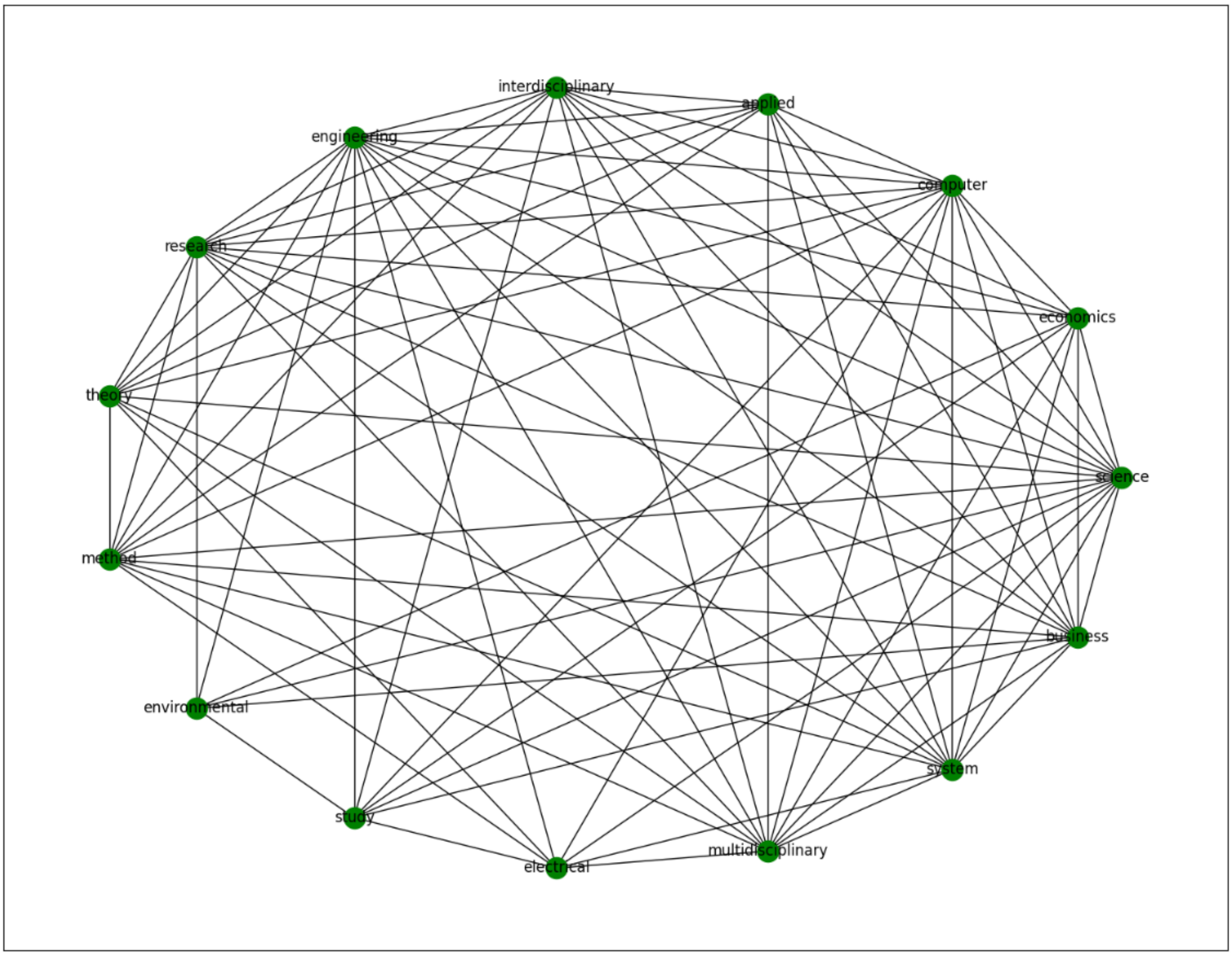


Figure 9. Visualization of Academic data Keywords Unigrams (own research)

Academic data Keywords Bigrams:

15 top nodes based on degree:

[('computer science', 103), ('engineering electrical', 49), ('science interdisciplinary', 48), ('science theory', 45), ('engineering multidisciplinary', 42), ('engineering civil', 39), ('science technology', 37), ('electrical electronic', 36), ('science information', 36), ('energy fuel', 34), ('theory method', 34), ('interdisciplinary application', 33), ('green sustainable', 32), ('sustainable science', 32), ('operation research', 31)].

The detailed visualization of the network is presented in Fig. 10.

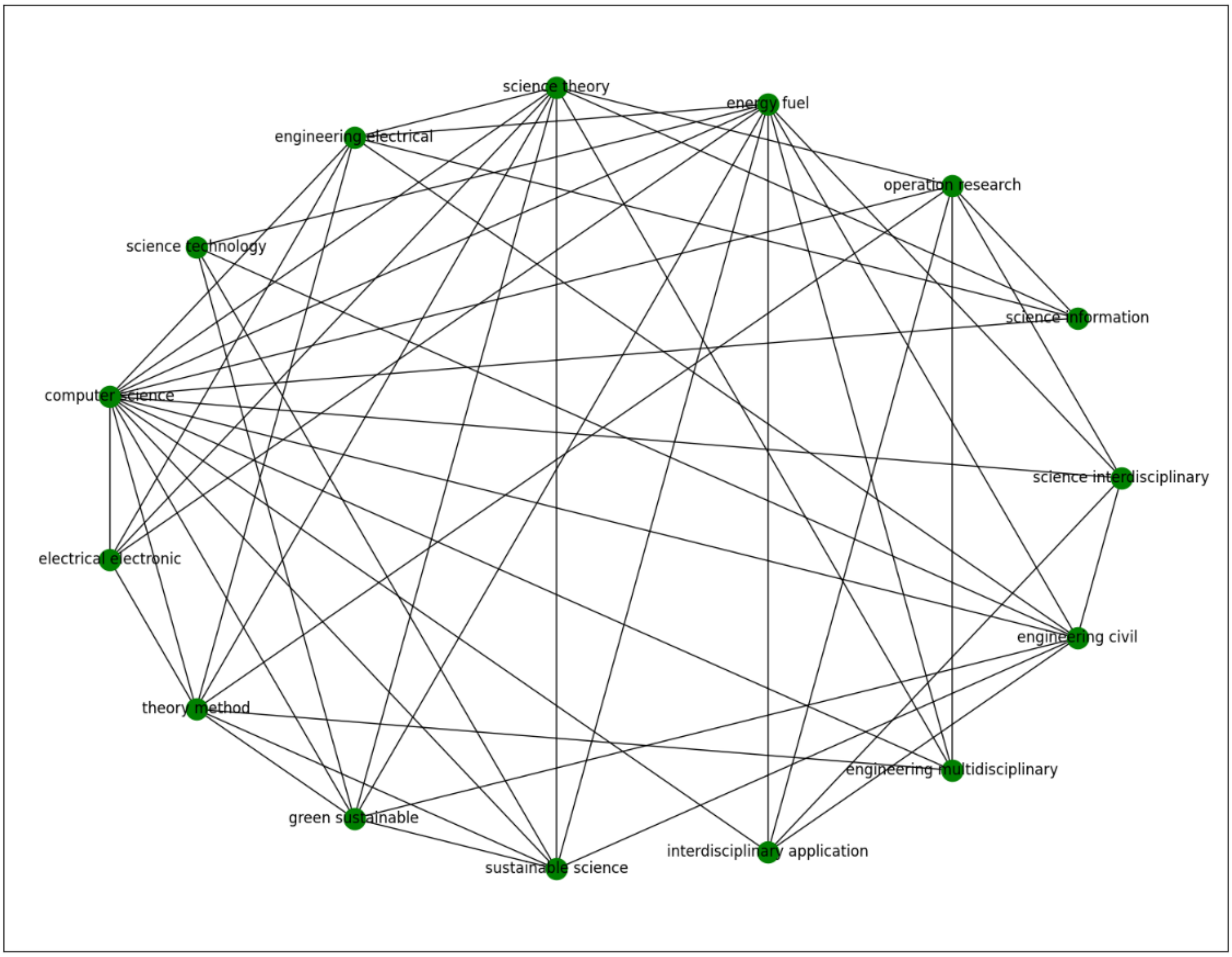


Figure 10. Visualization of Academic data Keywords Bigrams (own research)

Academic data Keywords Trigrams:

15 top nodes based on degree:

[('computer science theory', 49), ('computer science information', 39), ('engineering electrical electronic', 36), ('science theory method', 34), ('green sustainable science', 31), ('science information system', 31), ('computer science interdisciplinary', 30), ('health care science', 28), ('science interdisciplinary application', 27), ('care science service', 24), ('sustainable science technology', 24), ('computer science artificial', 21), ('education educational research', 21), ('science artificial intelligence', 21), ('automation control system', 19)].

The detailed visualization of the network is presented in Fig. 11.

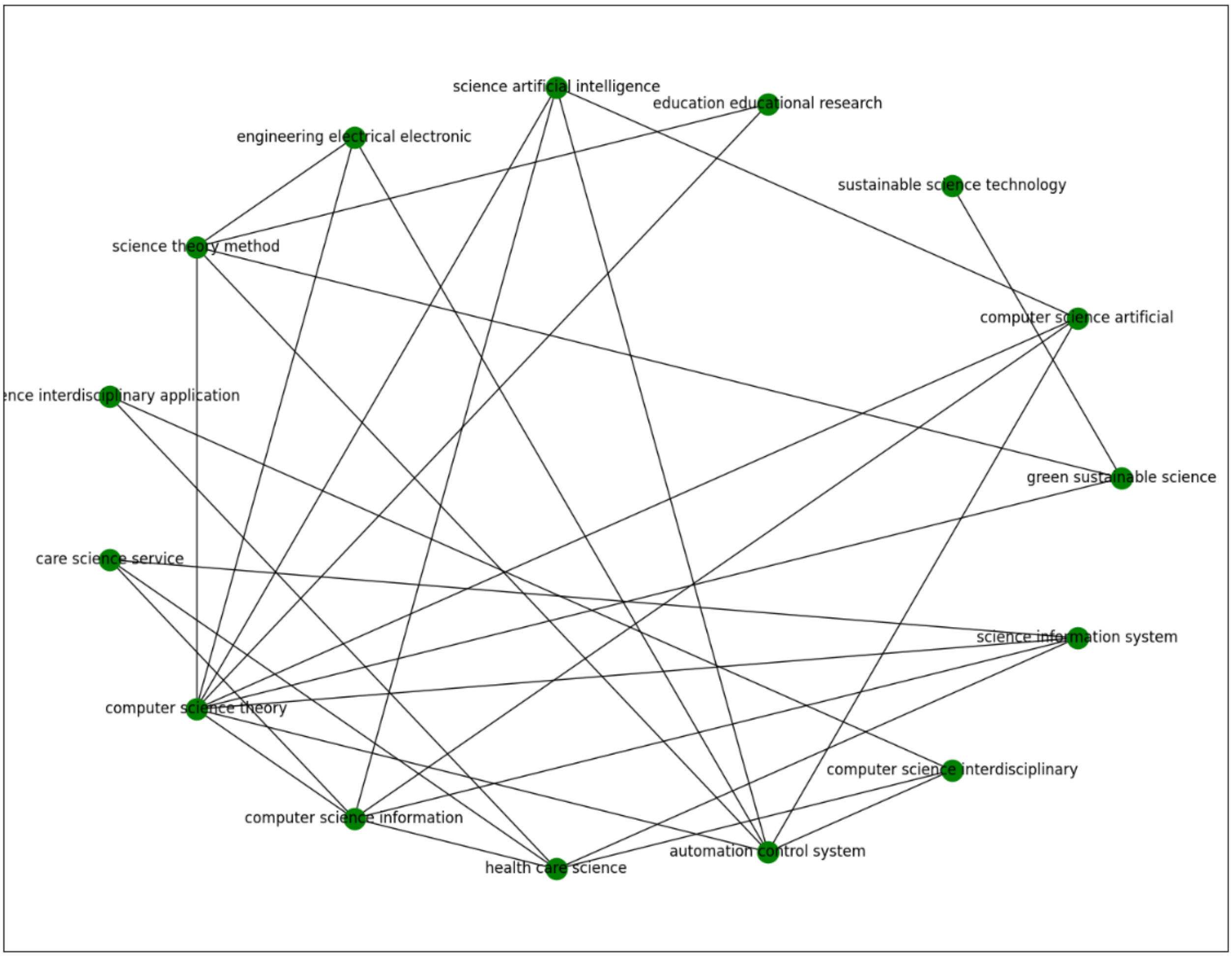


Figure 11. Visualization of Academic data Keywords Trigrams (own research)

Academic data Title Unigrams:

15 top nodes based on degree:

[('system', 436), ('software', 421), ('project', 392), ('model', 382), ('based', 379), ('approach', 369), ('construction', 362), ('using', 309), ('development', 307), ('agile', 306), ('research', 294), ('study', 288), ('practice', 280), ('managing', 278), ('analysis', 273)].

The detailed visualization of the network is presented in Fig. 12.

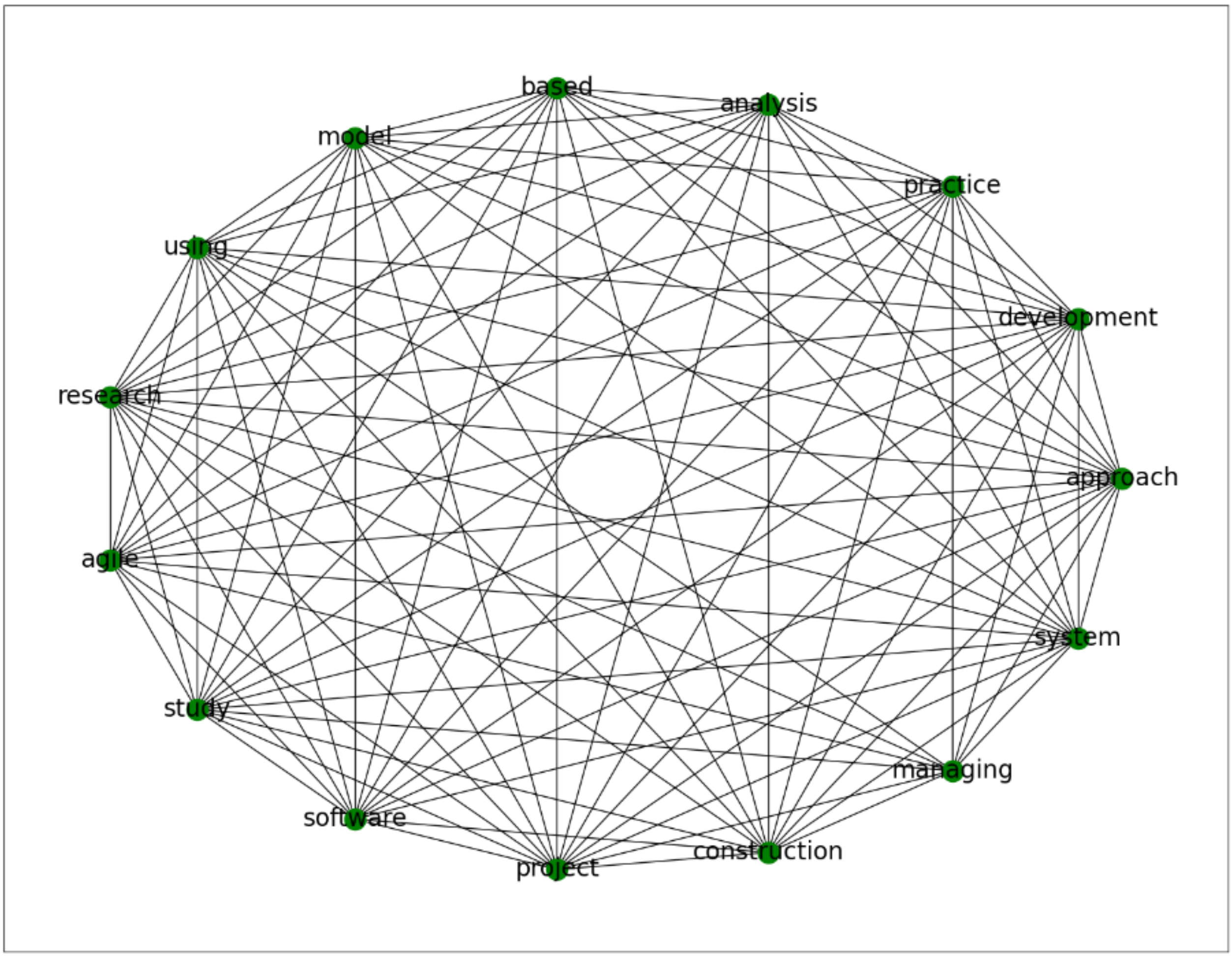


Figure 12. Visualization of Academic data Title Unigrams (own research)

Academic data Title Bigrams:

15 top nodes based on degree:

[('case study', 221), ('managing project', 120), ('information system', 109), ('decision making', 73), ('critical chain', 66), ('software development', 53), ('based learning', 43), ('best practice', 43), ('product development', 43), ('success factor', 43), ('web based', 38), ('based system', 35), ('risk assessment', 35), ('special issue', 35), ('decision support', 33)].

The detailed visualization of the network is presented in Fig. 13.

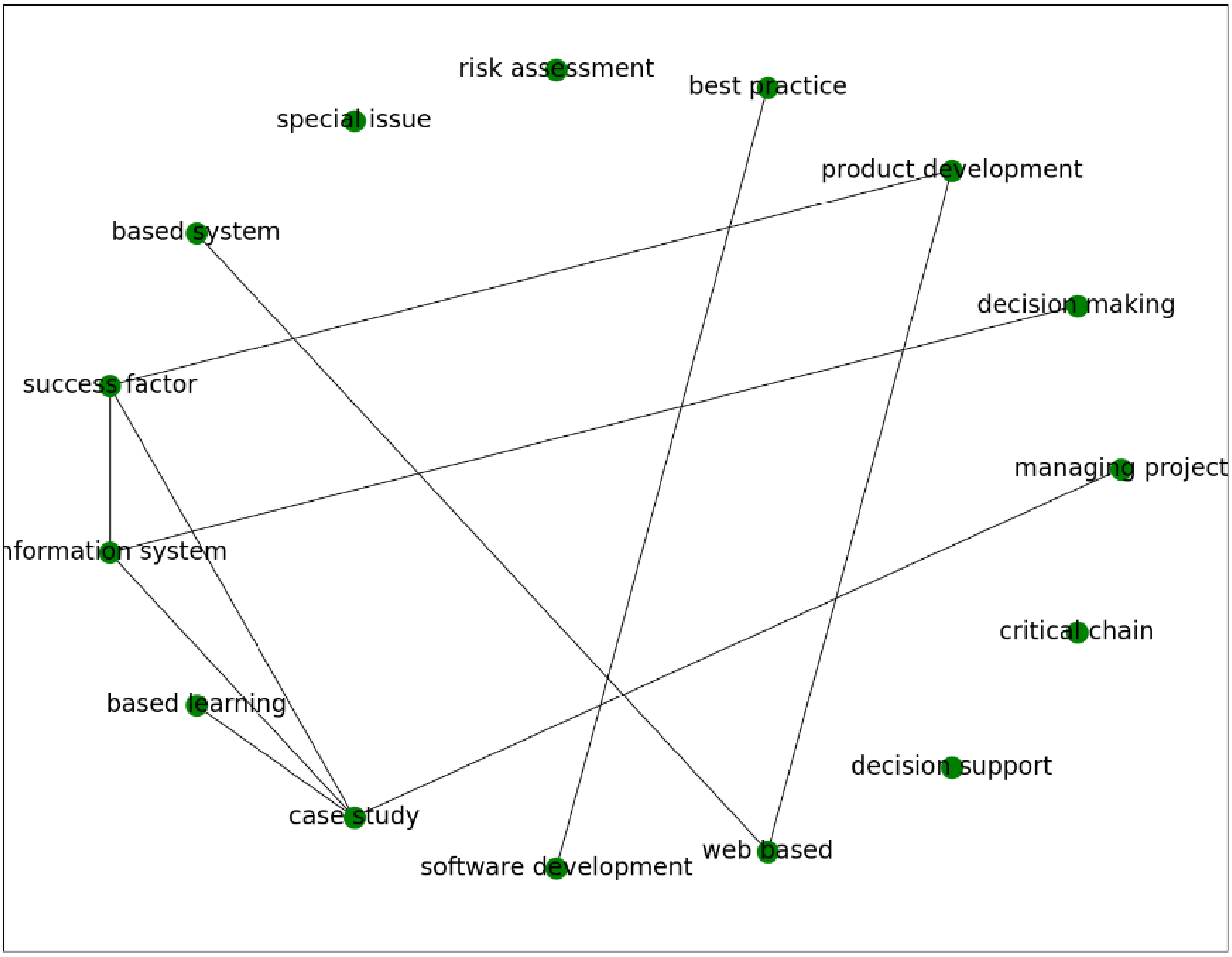


Figure 13. Visualization of Academic data Title Bigrams (own research)

Academic data Title Trigrams:

15 top nodes based on degree:

[('critical success factor', 22), ('global software development', 18), ('small medium sized', 15), ('decision support system', 14), ('new product development', 14), ('fuzzy multi objective', 11), ('web based system', 11), ('gert type network', 10), ('robust time cost', 10), ('social network analysis', 10), ('time cost tradeoff', 10), ('open source tool', 9), ('agile stage gate', 8), ('analytic hierarchy process', 8), ('green supply chain', 8)].

The detailed visualization of the network is presented in Fig. 14.

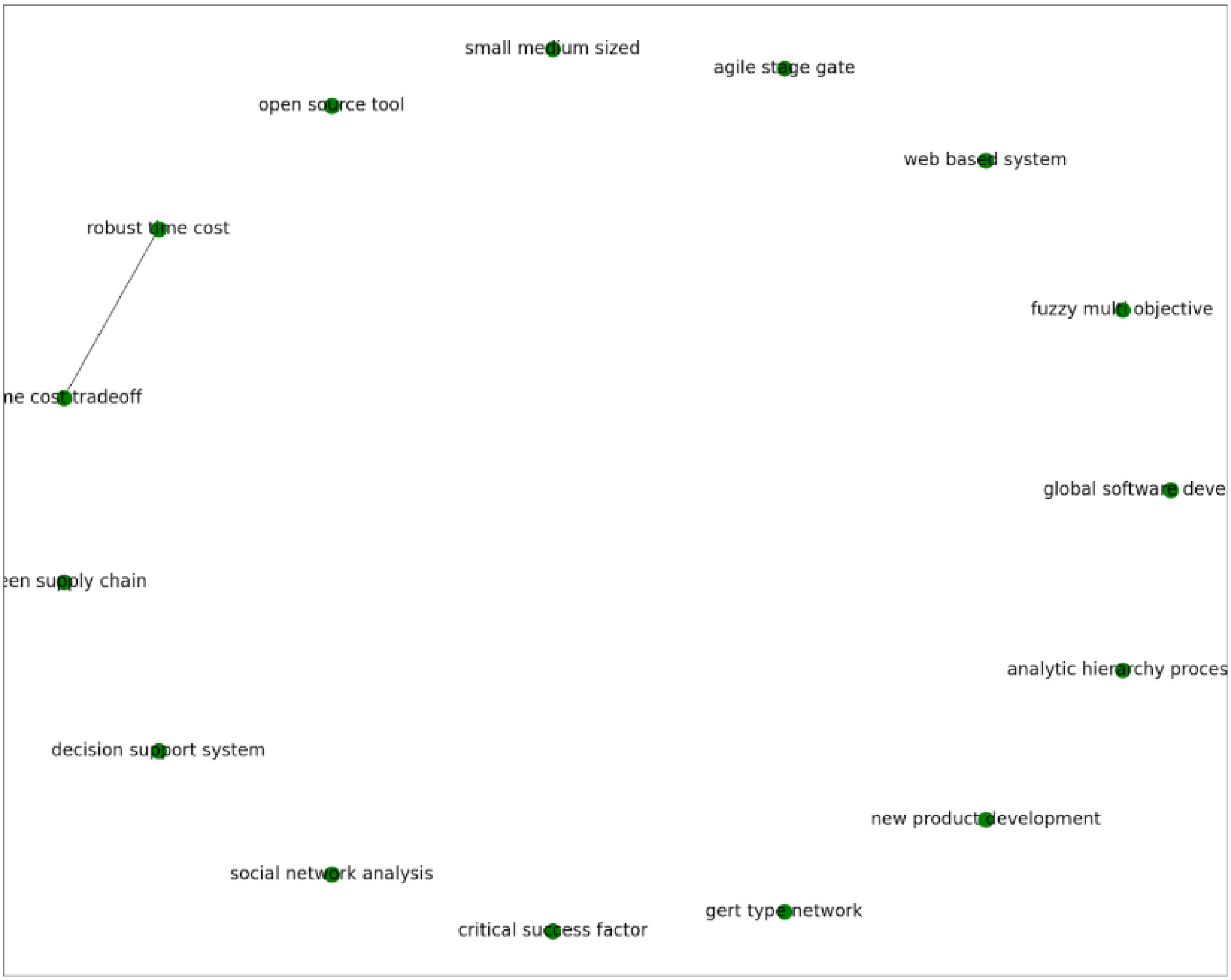


Figure 14. Visualization of Academic data Title Trigrams (own research)

Academic data Abstract Unigrams:

15 top nodes based on degree:

[('project', 8282), ('study', 7794), ('paper', 7693), ('research', 7354), ('based', 7348), ('result', 7225), ('process', 6871), ('development', 6660), ('approach', 6637), ('analysis', 6213), ('method', 6184), ('model', 6077), ('used', 6018), ('using', 5953), ('system', 5952)].

The detailed visualization of the network is presented in Fig. 15.

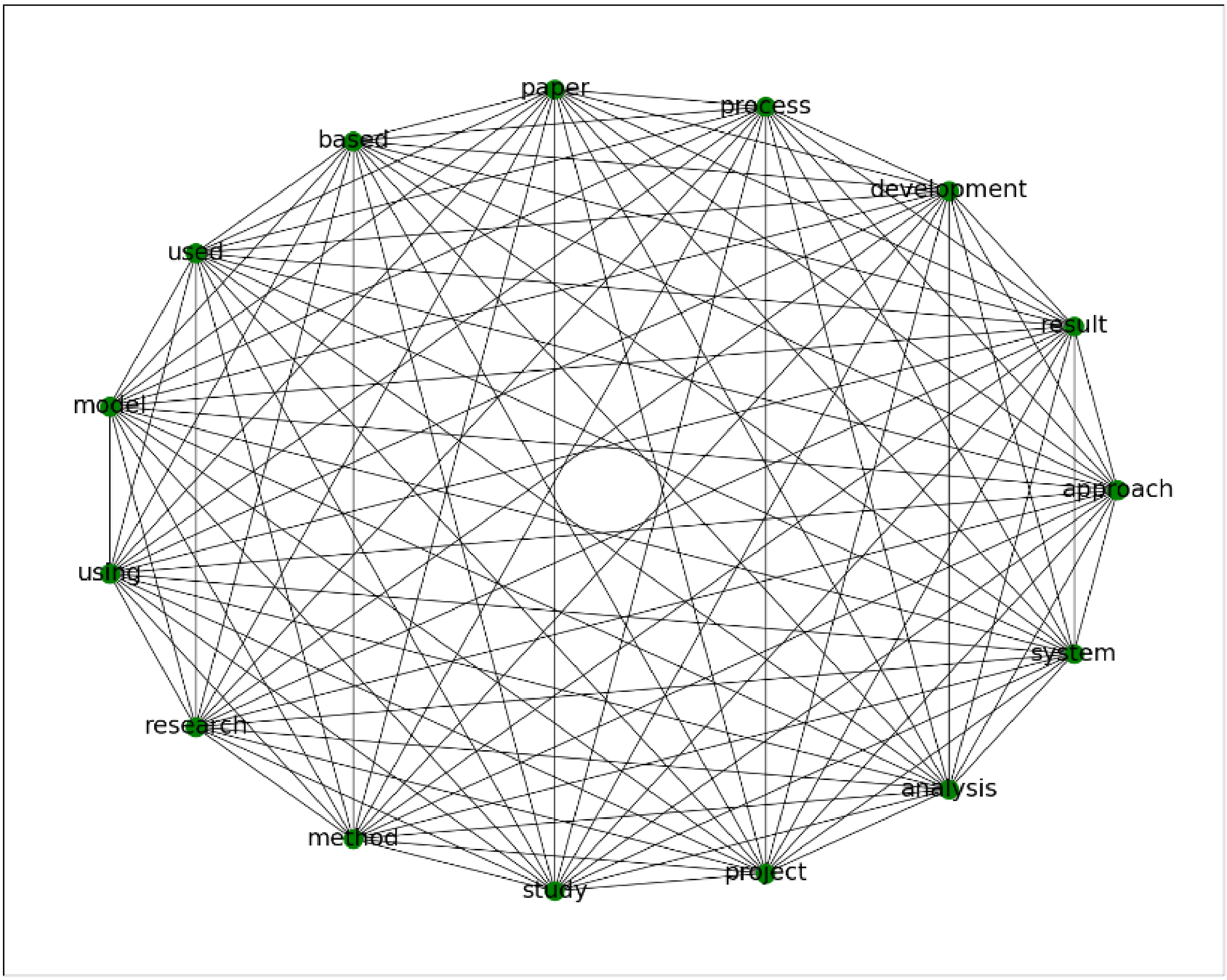


Figure 15. Visualization of Academic data Abstract Unigrams (own research)

Academic data Abstract Bigrams:

15 top nodes based on degree:

[('case study', 23704), ('right reserved', 19308), ('elsevier ltd', 15233), ('paper present', 14607), ('construction industry', 13266), ('decision making', 13192), ('literature review', 11918), ('result show', 11636), ('design methodology', 11430), ('methodology approach', 11207), ('construction project', 10971), ('originality value', 10332), ('life cycle', 9827), ('ipma right', 9419), ('software development', 9033)].

The detailed visualization of the network is presented in Fig. 16.

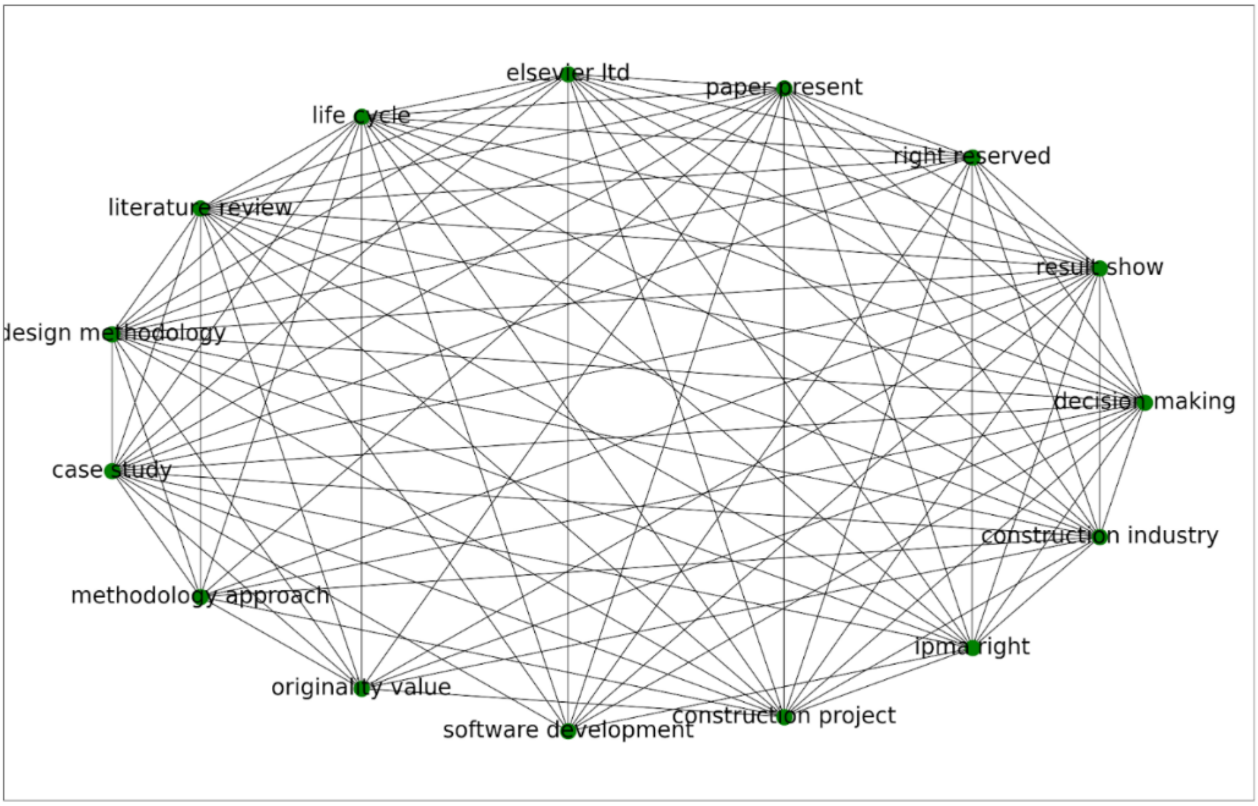


Figure 16. Visualization of Academic data Abstract Bigrams (own research)

Non-Academic data Title Unigrams:

15 top nodes based on degree:

[('manager', 152), ('job', 129), ('career', 33), ('best', 24), ('uk', 23), ('senior', 22), ('project', 20), ('com', 19), ('software', 19), ('sr', 18), ('may', 17), ('tool', 15), ('agile', 12), ('graduate', 12), ('digital', 11)].

The detailed visualization of the network is presented in Fig. 17.

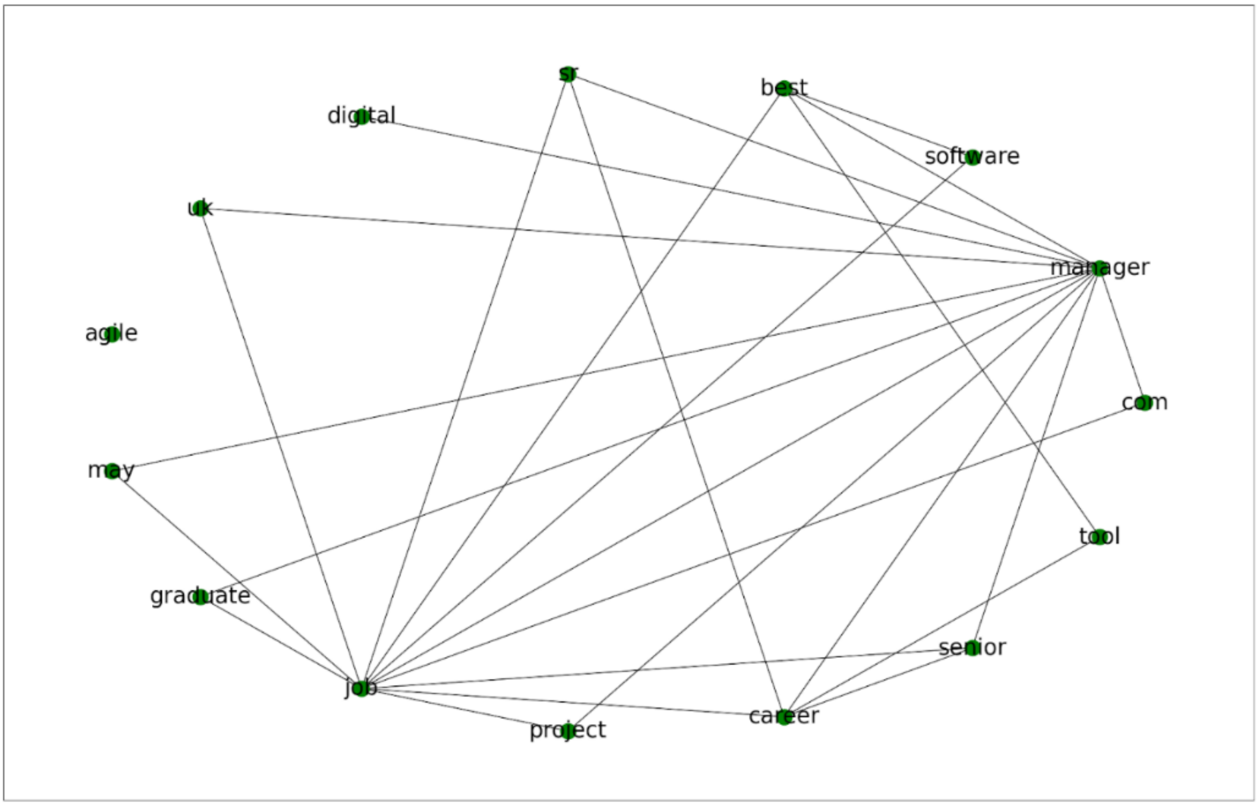


Fig 17. Visualization of Non-Academic data Title Unigrams (own research)

Non-Academic data Title Bigrams:

15 top nodes based on degree:

[('manager job', 51), ('co uk', 10), ('change job', 9), ('senior manager', 9), ('certification training', 8), ('made easy', 8), ('pmp certification', 8), ('advisor irvine', 5), ('arlington va', 5), ('authority wmca', 5), ('ca selectleaders', 5), ('capital work', 5), ('cbis senior', 5), ('combined authority', 5), ('course online', 5)].

The detailed visualization of the network is presented in Fig. 18.

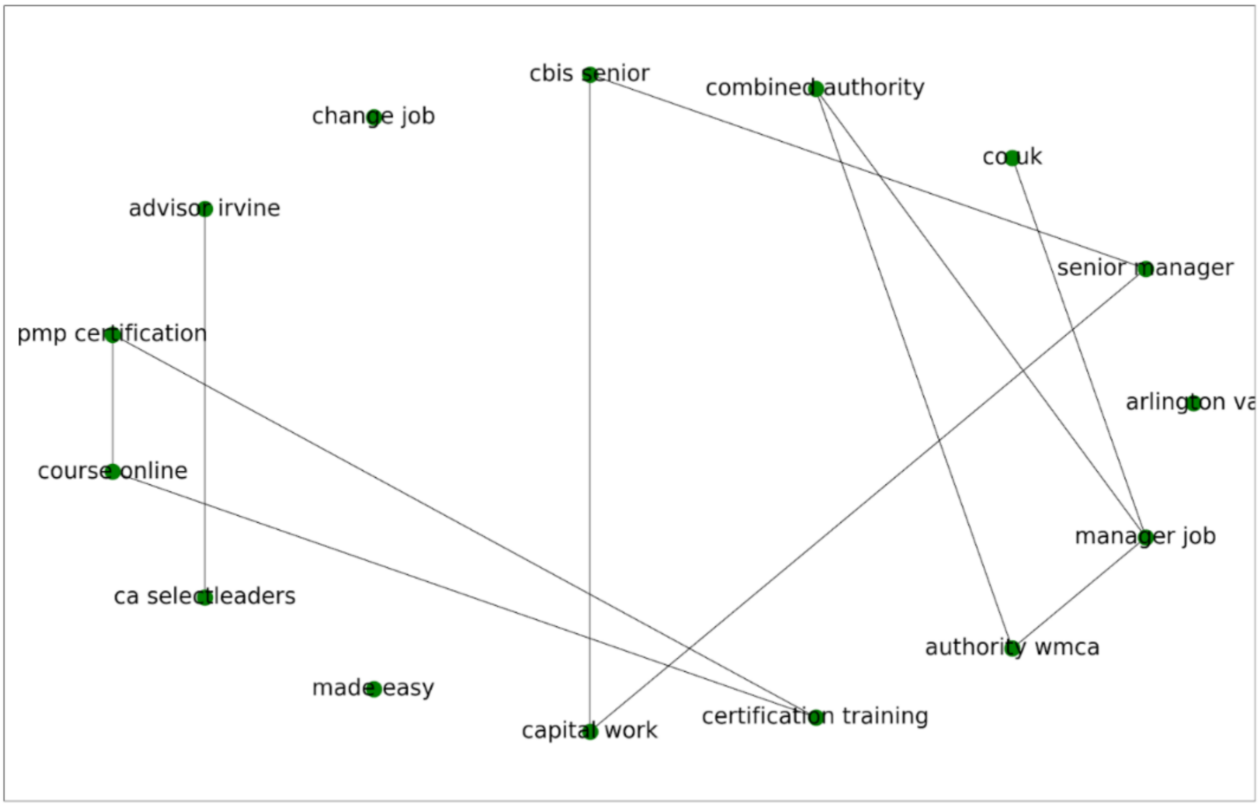


Fig 18. Visualization of Non-Academic data Title Bigrams (own research)

Non-Academic data Title Trigrams:

15 top nodes based on degree:

[('pmp certification training', 7), ('advisor irvine ca', 4), ('capital work csiro', 4), ('cbis senior manager', 4), ('certification training course', 4), ('change job ireland', 4), ('combined authority wmca', 4), ('course online may', 4), ('digital market arlington', 4), ('engineering pretoria east', 4), ('gartner digital market', 4), ('houston texas united', 4), ('ireland morgan mckinley', 4), ('irvine ca selectleaders', 4), ('isle wight reed', 4)].

The detailed visualization of the network is presented in Fig. 19.

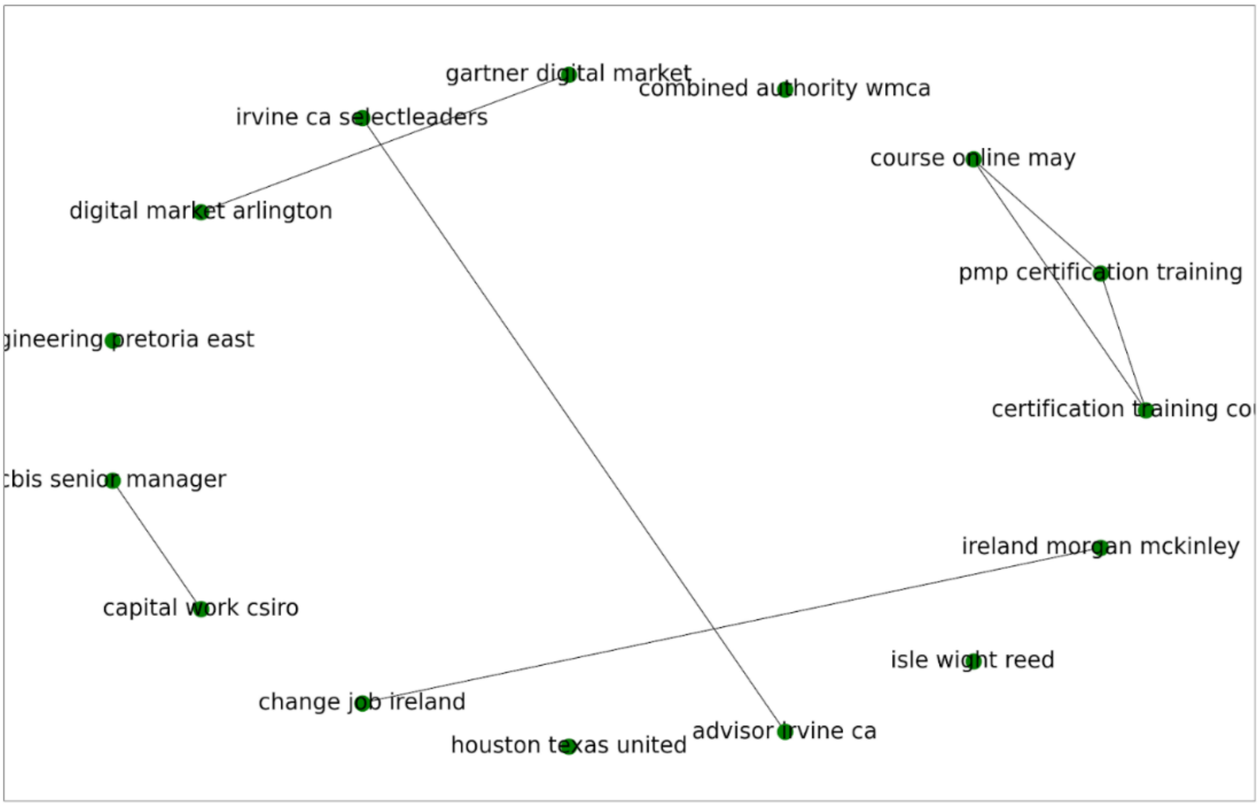


Fig 19. Visualization of Non-Academic data Title Trigrams (own research)

Non-Academic data Body Unigrams:

15 top nodes based on degree:

[('time', 6932), ('work', 6770), ('project', 6735), ('team', 6731), ('service', 6583), ('need', 6444), ('business', 6368), ('manager', 6312), ('make', 6246), ('data', 6213), ('new', 6207), ('use', 6166), ('process', 6141), ('also', 6137), ('provide', 6113)].

The detailed visualization of the network is presented in Fig. 20.

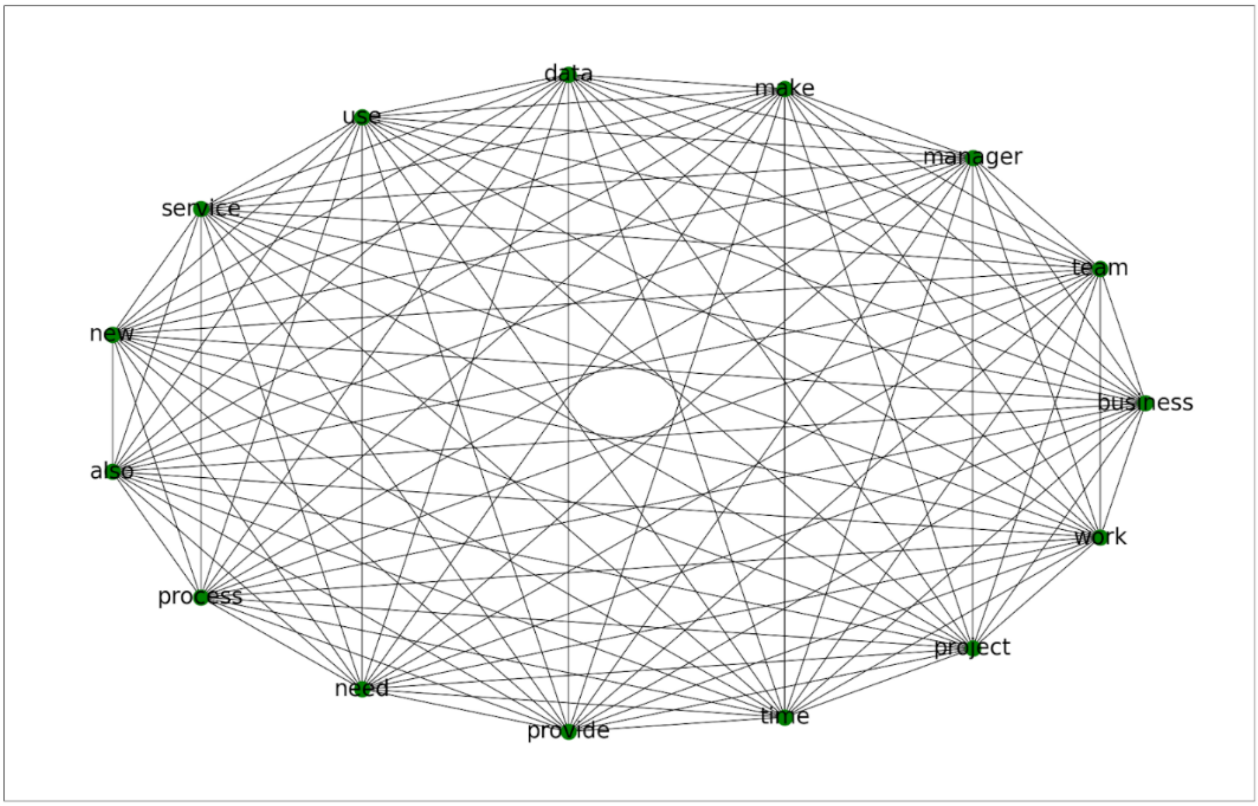


Figure 20. Visualization of Non-Academic data Body Unigrams (own research)

4.5. Network analysis

Upon obtaining network and graph visualizations, the next step is to engage in a comprehensive analysis of these graphs in order to extract valuable insights from the underlying networks. The initial feedback provided by network visualization sets the foundation for subsequent analytical processes. Among the various analytical techniques available, one of the most commonly employed methods is the examination of node degree, which measures the frequency and significance of nodes within the network.

Node degree analysis holds a paramount position in network analysis due to its ability to reveal critical information about the importance of individual nodes. By quantifying the number of times a particular node is encountered within the network, this analysis offers valuable insights into the node's prominence and relevance within the overall structure. Essentially, node degree analysis serves as a primary and essential tool for comprehending network dynamics and the significance of specific nodes within it. Through this analysis, I gain a conceptual understanding of a node's influence, which can guide further exploration and decision-making in network-related endeavors.

Network analysis encompasses a rich array of algorithmic tools, each tailored to examine distinct facets of complex networks. Two primary branches of these algorithms are node analysis and edge analysis. The choice among these analytical strategies largely depends on the specific goals and features of the task being examined.

In the current study, I place a particular emphasis on nodes within the network due to their pivotal role in shaping network behavior. Consequently, I have predominantly employed node analysis algorithms to explore and highlight the most influential nodes within the network.

Node Analysis algorithms center their scrutiny on individual nodes within the network. They seek to quantify and assess various attributes associated with nodes, including but not limited to centrality measures (e.g., degree, betweenness, closeness), influence, or hierarchical positioning. Node analysis is particularly well-suited for tasks where identifying key actors or entities is imperative. For instance, in a social network, node analysis may help pinpoint influential individuals, while in a transportation network, it could reveal critical junctions or endpoints.

Conversely, edge analysis algorithms shift the analytical lens towards the connections or edges that link nodes within the network. These algorithms aim to decipher the intricate relationships between connected nodes, investigating aspects like the strength, direction, and patterns of interactions. The examination of edges is vital for comprehending dynamic phenomena like the dissemination of information, chain reactions, or potential weak points in a network's framework.

In my study, the rationale behind the preeminence of node analysis is deeply rooted in my research objectives and the inherent characteristics of the network under investigation. By employing node analysis algorithms extensively, I seek to uncover and illuminate the central nodes that wield substantial influence, occupy key positions, or serve as linchpins in the network's overall structure. This strategic choice is motivated by the recognition that individual nodes often serve as critical hubs or influencers in network dynamics.

This analytical approach aligns seamlessly with my research goals, allowing us to delve deeper into the network's intricate architecture and functionality. Moreover, it enriches the broader body of knowledge in the field of network analysis by shedding light on the fundamental role of nodes in shaping complex systems.

I rigorously applied a diverse array of node analysis methodologies to comprehensively investigate and scrutinize the underlying characteristics and behaviors of the nodes within my network framework. Within my project, I executed node analysis algorithms as follows:

1. *Degree Centrality:*
    - Definition: Degree centrality measures the number of edges connected to a node in a network.
    - Mathematical formula: For an undirected graph, the degree centrality of a node is calculated by dividing the number of edges connected to the node by the total number of nodes in the graph minus one. For a directed graph, the in-degree and out-degree centrality can be calculated similarly.
    - Outcome: The degree centrality score represents the importance or influence of a node based on its connectivity. Nodes with higher degree centrality values are considered more central in the network.

2. *Betweenness Centrality:*
    - Definition: Betweenness centrality identifies nodes that act as bridges or intermediaries between other nodes in a network.
    - Mathematical formula: Betweenness centrality is computed by tallying the number of shortest paths passing through a node divided by the total number of shortest paths between all pairs of nodes.
    - Outcome: The betweenness centrality score quantifies the extent to which a node controls the flow of information in the network. Nodes with higher betweenness centrality values play a critical role in maintaining communication between different parts of the network.

3. *Closeness Centrality:*
   - Definition: Closeness centrality measures how quickly a node can reach all other nodes in a network.
   - Mathematical formula: Closeness centrality is calculated as the reciprocal of the average shortest path length from a node to all other reachable nodes in the graph.
   - Outcome: The measure of closeness centrality quantifies the proximity of a node to all other nodes in the network, considering the shortest path length. Nodes that exhibit higher values of closeness centrality are more reachable and maintain a shorter mean distance to the rest of the nodes in the network.
4. *Eigenvector Centrality:*
   - Definition: Eigenvector centrality measures the influence or importance of a node in a network based on the centrality of its neighboring nodes.
   - Mathematical formula: Eigenvector centrality assigns a score to each node proportional to the sum of the eigenvector centralities of its neighbors. The centrality values are obtained by solving the eigenvector equation.
   - Outcome: Eigenvector centrality identifies nodes that are not only well-connected but also connected to other highly central nodes. Nodes with higher eigenvector centrality scores are considered more influential within the network.
5. *PageRank Centrality:*
   - Definition: PageRank centrality is a variant of eigenvector centrality and measures the importance of a node in a network based on the principle of "voting" or "recommendation" from other important nodes.
   - Mathematical formula: PageRank centrality uses a recursive formula that assigns a score to each node based on the sum of the PageRank scores of its incoming neighbors, taking into account the importance (PageRank) of those neighbors as well.
   - Outcome: PageRank centrality is commonly used to rank web pages in search engine algorithms. In networks, it identifies nodes that are highly recommended or endorsed by other important nodes. Nodes with higher PageRank centrality values are considered more influential or significant within the network structure.
6. *HITS (Hubs and Authorities):*
   - Definition: HITS algorithm identifies two kinds of important nodes in a network: hubs and authorities. Hubs are nodes that connect to many high-authority nodes, while authorities are nodes that are connected to by many high-hub nodes.
   - Mathematical formula: HITS computes two scores for each node - the hub score and the authority score. The hub score of a node is computed as the sum of the authority scores of its outgoing neighbors, and the authority score of a node is computed as the sum of the hub scores of its incoming neighbors. The scores are then normalized iteratively until convergence.
   - Outcome: HITS algorithm provides a way to identify nodes that are both influential and connected to other influential nodes. Nodes with higher hub scores are considered as hubs, while nodes with higher authority scores are considered as authorities.
7. *Clustering Coefficient:*
   - Definition: The clustering coefficient measures the degree to which nodes in a graph tend to cluster together.
   - Mathematical formula: The clustering coefficient of a node is calculated as the fraction of triangles (closed loops of three nodes) that the node participates in, divided by the total number of possible triangles centered on that node.

- Outcome: The clustering coefficient provides insight into the local connectivity patterns around each node. Nodes with higher clustering coefficients indicate the presence of tightly connected groups or communities within the network.

8. *Assortativity Coefficient:*
   - Definition: The assortativity coefficient quantifies the degree to which nodes in a network are likely to link with other nodes that possess analogous characteristics.
   - Mathematical formula: The assortativity coefficient is computed by comparing the similarity of node attributes (e.g., degree, age, etc.) between pairs of connected nodes in the network.
   - Outcome: The assortativity coefficient provides insights into the assortative mixing patterns in a network. Positive values indicate that nodes preferentially connect to others with similar attributes, while negative values suggest disassortative mixing where nodes connect to nodes with dissimilar attributes.

9. *Load Centrality:*
   - Definition: : Load centrality quantifies the significance or impact of a node, taking into account the volume of traffic it handles within a network.
   - Mathematical formula: Load centrality allocates a score to each node in relation to the aggregate number of shortest paths within the network that traverse that particular node.
   - Outcome: Load centrality identifies nodes that act as critical points of congestion or traffic flow within the network. Nodes with higher load centrality values carry a larger proportion of the overall network traffic.

10. *Coreness Centrality:*
    - Definition: centrality evaluates the structural significance of a node, considering its level in the k-core breakdown of the network.
    - Mathematical formula: Coreness centrality assigns a score to each node based on the maximum value of k for which the node belongs to the k-core of the network. The k-core represents a subgraph where each node is connected to at least k nodes within the subgraph.
    - Outcome: Coreness centrality identifies nodes that are part of the densest and most central regions of the network. Nodes with higher coreness centrality values are considered more structurally important within the network.

11. *Harmonic Centrality:*
    - Definition: Harmonic centrality measures the average reciprocal of the shortest path distances from a node to all other nodes in the network.
    - Mathematical formula: Harmonic centrality of a node is calculated as the sum of the reciprocals of the shortest path distances from that node to all other nodes in the network.
    - Outcome: Harmonic centrality identifies nodes that are well-connected to other nodes and have short paths to reach most other nodes. Nodes with higher harmonic centrality values are considered more central within the network.

12. *Average Neighbor Degree:*
    - Definition: Average neighbor degree measures the average degree of the neighbors of a node in the network.
    - Mathematical formula: Average neighbor degree of a node is calculated as the average of the degrees of all its neighboring nodes.
    - Outcome: Average neighbor degree helps identify nodes whose neighbors tend to have higher or lower degrees. Nodes with higher average neighbor degrees are considered more central within their local neighborhoods.

13. *Closeness Vitality:*

- Definition: Closeness Vitality is a measure that quantifies the impact of node removal on the overall network closeness centrality. It assesses how the removal of nodes affects the accessibility and efficiency of communication within the network.
- Mathematical Formula: Closeness Vitality is calculated by summing up the absolute differences in the closeness centrality values of all nodes before and after the removal of each node.
- Outcome: Closeness Vitality helps identify critical nodes whose removal would significantly decrease the overall closeness centrality of a network. It highlights nodes that play vital roles in maintaining efficient communication pathways and are crucial for ensuring short path lengths between nodes in the network.

It must be mentioned that my analytical process entailed the consolidation of all relevant analyses for each individual time period into a unified file. This consolidation approach facilitated a comprehensive perspective, as it enabled us to base my analyses on a consolidated dataset encompassing all n-grams, including unigrams, bigrams, and trigrams.

This methodological decision to aggregate all analytical data into a single file was deliberate and well-considered. It allowed us to work with a holistic dataset, thereby ensuring that my analyses were grounded in a comprehensive representation of the linguistic components under scrutiny. Consequently, this approach enabled us to derive insights that transcended the confines of individual n-grams or periods, affording us a richer and more nuanced understanding of the underlying patterns and dynamics within my dataset.

It is essential to underscore that the datasets assembled for this research exhibit a pronounced focus on the theme of 'project management.' Consequently, it becomes evident that the keywords 'project,' 'management,' and 'project management' are likely to play a pivotal role in the subsequent network analysis. To mitigate the potential bias introduced by these prevalent terms, I undertook the task of eliminating them from the dataset before subjecting it to my analysis algorithms. The presence of these frequently occurring words within the network could significantly skew the outcomes by overshadowing the contributions of other nodes. Furthermore, the stark contrast in the analytical results between these key terms and other words within the dataset can render the interpretation of findings intricate and, at times, elusive.

Moreover, the distinctiveness in the analytical outcomes between these focal keywords and other less frequent terms underscores the potential challenges in drawing meaningful inferences from the data. The presence of such significant disparities could necessitate a more careful and nuanced interpretation of the findings, taking into account the potential confounding effects introduced by these dominant keywords. This complexity underscores the importance of my effort to untangle the web of interconnected terms within the dataset, ensuring a more comprehensive and accurate analysis.

Furthermore, it is imperative to acknowledge that the aforementioned analysis techniques represent the conventional and widely adopted methods employed in network analysis across various domains. These established methodologies serve as the cornerstone of network analysis tasks due to their extensive utilization and proven effectiveness.

In addition to these conventional approaches, my research venture has ventured into uncharted territory by implementing unconventional analysis techniques that deviate from the norm. Despite their non-conformity to the traditional standards, these novel methods have exhibited promising results, thus opening up new avenues for exploration and innovation in the field of network analysis. This diversification in analytical strategies not only enriches the existing knowledge base but also highlights the potential for discovering groundbreaking insights and solutions within the realm of network analysis.

I have introduced two distinct analytical approaches, although it is important to emphasize that these analyses are not entirely independent and have been derived from broader analytical frameworks. These two approaches are the "Sum of Analysis" and the "Weighted Sum of Analysis."

In the "Sum of Analysis" method, I aggregate all the individual analyses mentioned in the preceding network analysis list. This straightforward summation provides a holistic view of the data, showcasing the collective insights gained from these individual analyses. It serves as a foundational step in understanding the data's overall characteristics and trends.

The "Weighted Sum of Analysis" approach, on the other hand, employs a more sophisticated method. Here, instead of simply summing the analyses, I assign each analysis a specific weight before aggregation. These weights are determined through Principal Component Analysis (PCA), a statistical technique used to identify the most influential components within a dataset.

The rationale behind incorporating PCA lies in its ability to provide additional insights into the data. By assigning weights based on PCA results, I can highlight the relative importance of each analysis in the broader context. This approach goes beyond a simple summation, allowing us to discern the significance of individual analyses and their impact on subsequent analyses.

In essence, the inclusion of PCA in the "Weighted Sum of Analysis" method offers a deeper level of understanding. It not only quantifies the contributions of each analysis but also reveals the hidden relationships and dependencies among them. This, in turn, enhances my ability to gain valuable insights from the successive analytical steps.

4.6. Thomas Kuhn science philosophy:

Thomas Kuhn's philosophy of science argues that scientific advancement is not characterized by a linear and continuous progression but rather by periodic "paradigm shifts." These shifts represent fundamental changes in the basic concepts and experimental practices of a scientific discipline, often triggered by an accumulation of anomalies that the current paradigm cannot account for. During periods of "normal science," the research conducted by the scientific community is based on an established paradigm—a framework that includes the beliefs, techniques, and standards shared by the community. This period is marked by cumulative progress within the confines of the existing paradigm. In contrast, "revolutionary science" emerges when the current paradigm is no longer sufficient to explain observed phenomena, leading to a scientific revolution and the establishment of a new paradigm.

Kuhn's seminal work, "The Structure of Scientific Revolutions," emphasizes that the notion of scientific truth at any given time is shaped not only by objective criteria but also by the prevailing consensus within the scientific community. This introduces a social dimension to the concept of scientific truth, suggesting that it is, to some extent, constructed by the community's subjective agreement rather than by empirical evidence alone.

Moreover, Kuhn's influence extends beyond the philosophy of science. The terms he introduced, such as "paradigm shift" and "normal science," have permeated academic and public discourse, illustrating the broad impact of his ideas. His philosophy underscores the importance of historical and social contexts in understanding the evolution of scientific knowledge, thereby challenging the simplistic view of science as a straightforward march towards truth.

4.7. I-indicator C- indicator and ICCO ranking:

This part examines the utility of various network analysis metrics, with a special focus on the I-indicator and the clustering coefficient. The I-indicator is highlighted as a significant tool for measuring a node's growth potential within a network by considering the ratio of its betweenness centrality to its degree. It is especially useful for identifying nodes with high growth potential, ranking nodes in terms of their influence, optimizing network interventions, and enhancing predictive models. This versatility allows it to be adapted across different fields.

The clustering coefficient, or C-indicator, measures the extent to which nodes in a network cluster together. It is crucial for understanding the local network structure, identifying community structures within networks, assessing network robustness, and guiding network optimization efforts. It is applicable in diverse areas such as social sciences, biology, and sociology.

Additionally, the thesis introduces the ICCO rank, which combines the I-indicator and C-indicator with the frequency of a keyword associated with a node. This rank offers a comprehensive view of a node's growth potential and significance within the network, with lower values indicating higher potential.

4.8. Relation between, Kuhn's philosophy and indicators:

Integrating Thomas Kuhn's philosophy of science with the application of I-indicator, C-indicator, and ICCO ranking involves understanding the stages of scientific development as Kuhn described them and then considering how these indicators might play a role in identifying and facilitating paradigm shifts.

Here's how Kuhn's stages of science could be understood in conjunction with these indicators:

- Pre-Paradigm Phase:

Before science reaches a paradigm, various competing schools of thought exist. In this phase, there is no consensus on the theories or methods that should be used. During this period, the I-indicator, C-indicator, and ICCO ranking could serve as analytical tools to assess the potential of various theories. For instance, the I-indicator could be used to evaluate the growth potential of different theories or methods within the scientific community, while the C-indicator could assess how well-integrated or connected a theory is within the existing network of scientific knowledge.

- Normal Science:

During normal science, the community operates under a consensus on what constitutes the paradigm—the set of practices and standards that define legitimate work in the field. The I-indicator and C-indicator could be used to monitor the standard research practice within the paradigm, identifying which theories or methods are central (high betweenness centrality and degree) and well-integrated (high clustering coefficient). The ICCO ranking could prioritize areas of research that are ripe for development, based on the frequency of key terms associated with emerging anomalies or questions that the current paradigm cannot answer.

- Crisis:

When the paradigm encounters enough anomalies that it cannot explain, a crisis occurs. Researchers might use the I-indicator to identify which alternative theories are gaining traction (showing growth potential) and the C-indicator to see which new theories are beginning to form their own clusters of consensus or are robust against the anomalies that challenge the old paradigm. The ICCO ranking would be particularly useful here, as it could help in determining which alternative theories are most relevant and robust, combining the structural position of theories (I and C indicators) with their contextual relevance (keyword occurrence).

- Revolutionary Science (Paradigm Shift):

If a crisis leads to a paradigm shift, a new framework emerges that better explains the data and resolves the anomalies. The I-indicator, C-indicator, and ICCO ranking can then be applied to the new paradigm to assess its growth potential, the integration of its theories within the scientific community, and the relevance of its key concepts and methods.

- Post-Revolution Science:

After a paradigm shift, normal science resumes under the new framework. The new paradigm is now subject to the same kinds of analyses that the old one was, using the I-indicator and C-indicator to ensure that the new theories and methods are solidifying their position within the scientific community, and using the ICCO ranking to continue refining and expanding the paradigm as new data and ideas emerge.

In summary, the I-indicator, C-indicator, and ICCO ranking could be seen as tools to quantify and analyze the structure and dynamics of scientific knowledge, potentially playing a role in identifying the conditions that lead to a paradigm shift as described by Thomas Kuhn. These metrics would be especially useful for gauging the interconnectedness and potential growth of theories and practices within the scientific community at different stages of scientific development.

In addition to these two analyses, I used another one like I-indicator, I-ranking, C-ranking, C-indicator, ICCO, and ICCO-ranking.

The I-indicator represents a significant addition to my research, serving as a valuable metric for assessing a node's growth potential within a network. This innovative metric is derived from the ratio of a node's betweenness centrality to its node degree, allowing for a more comprehensive understanding of a node's role in the network's structure and dynamics(Abbasi, Altmann & Hossain 2011; AbdiKhalife, Dunay & Illés 2021).

The utility of the I-indicator in my research is multifaceted and can be particularly advantageous for several reasons(Kumar & Markscheffel 2016):

1. Growth Potential Identification: A primary benefit of the I-indicator is its capacity to pinpoint nodes that exhibit significant potential for expansion within a network. By considering both betweenness centrality and node degree, it offers a more nuanced perspective on a node's potential to exert influence and expand its reach in the network.

2. Enhanced Node Ranking: The I-ranking, which ranks nodes based on their I-indicator values, provides a systematic way to prioritize nodes within the network. This ranking can be instrumental in pinpointing the most influential or strategically significant nodes, streamlining the focus of my analyses.

3. Network Optimization: In scenarios where optimizing network resources or interventions is critical, understanding which nodes possess high growth potential becomes indispensable. The I-indicator facilitates the strategic allocation of resources to maximize their impact within the network.

4. Predictive Modeling: Incorporating the I-indicator into predictive models can enhance my ability to forecast the future behavior and growth trajectories of nodes within the network. This predictive capability has practical applications in fields such as social network analysis, transportation planning, and epidemiology, where anticipating network dynamics is crucial.

5. Adaptability Across Domains: The versatility of the I-indicator makes it a valuable tool in various domains, including social sciences, engineering, biology, and beyond. Researchers across these fields can leverage this concept to gain deeper insights into network structures, enabling them to make informed decisions and predictions specific to their respective areas of study.

The clustering coefficient Ci of node i is defined as:

$$C_i = \frac{2E_i}{D_i(D_i - 1)},$$

where Ei is the actual number of edges existing among the neighboring nodes of node i.

From a geometric point of view, Ci is also defined as:

$$C_i = \frac{number\ of\ complete\ triangles\ with\ corner\ i}{number\ of\ triangular\ graphs\ with\ corner\ i}.$$

The clustering coefficient, often denoted as C-indicator or simply C, is a fundamental concept in network theory and graph theory. It quantifies the extent to which nodes in a network have a tendency to aggregate or establish closely-connected clusters. This metric provides valuable insights into the structure and connectivity of a network and is widely used in various fields, including social sciences, biology, computer science, and sociology. Let's delve into its importance and how it can be useful(AbdiKhalife, Dunay & Illés 2021; Kumar & Markscheffel 2016).

Clustering coefficient helps us comprehend the local structure of a network. It quantifies the tendency of nodes to form triangles, indicating the extent to which neighbors of a node are connected to each other.

High clustering coefficients suggest that a network has a modular or community structure, where nodes are more densely connected within their groups than with nodes outside those groups.

In social networks, a high clustering coefficient can indicate the presence of tightly-knit friend groups or communities. Detecting such communities is crucial for understanding social dynamics, information diffusion, and targeted marketing. In biological networks, clustering coefficients help identify functional modules in protein-protein interaction networks or gene regulatory networks.

The clustering coefficient can be used to assess the robustness of a network. A network with a high clustering coefficient is often more resilient to random node failures because the remaining nodes are still likely to be well-connected. On the other hand, networks with low clustering coefficients may be more vulnerable to targeted attacks, as removing a few key nodes can disrupt the connectivity between different parts of the network(Kumar & Markscheffel 2016).

In recommendation systems, clustering coefficients can be used to find nodes with similar neighbors, which can aid in suggesting relevant items or connections to users. For example, in a social network, if two users have a high clustering coefficient, it suggests they share common friends or interests, making it more likely that they would be interested in each other's posts or recommendations.

Understanding the clustering coefficient can guide network optimization efforts. For instance, in communication networks, increasing clustering can lead to better fault tolerance and more efficient data transmission. In transportation networks, clustering can help identify regions where traffic congestion is likely to occur, enabling more targeted infrastructure improvements.

In sociology and psychology, the clustering coefficient can be used to study social networks and the formation of social bonds. It helps researchers analyze the strength and stability of interpersonal relationships within a group. High clustering coefficients in social networks may indicate a sense of close-knit communities and strong social cohesion (AbdiKhalife, Dunay & Illés 2021).

When assessing network nodes, a multifaceted approach is deployed to gauge their significance and potential for future growth. This process involves the consideration of two key indicators: the "I indicator" and the "cluster coefficient (C-indicator)." These indicators are systematically ranked for each node under examination. Subsequently, their respective rankings are enriched by factoring in an additional metric—the frequency of a specific keyword associated with the node. This amalgamated ranking methodology results in the emergence of a novel metric termed the "ICCO rank," an abbreviation derived from "I indicator, cluster coefficient (C-indicator), and occurrence ranking."

The ICCO rank represents a holistic assessment that not only encapsulates the inherent characteristics of a node, as signified by the I and C indicators, but also incorporates contextual information represented by the frequency of the keyword occurrence. This comprehensive ranking metric furnishes valuable insights into the potential for growth and significance of network nodes.

To decipher the ICCO rank, a ranking system is instituted, whereby nodes are arranged in ascending order based on their ICCO rank values. In this hierarchical arrangement, lower ICCO rank values denote a higher potential for future growth and significance within the network. Essentially, a node attaining an ICCO rank of 1 conveys the highest conceivable potential for advancement within the network (AbdiKhalife, Dunay & Illés 2021).

This analytical framework furnishes a structured and informative means of evaluating the growth prospects of network nodes. It empowers researchers to prioritize their attention toward nodes with the greatest potential to contribute to the advancement of their study. The ICCO rank, therefore, emerges as an invaluable tool in the realm of academic research and network analysis, enabling more discerning decision-making and resource allocation in research endeavors.

The fabric of scientific advancement, according to Thomas Kuhn's revolutionary philosophy, is characterized by episodic paradigm shifts rather than a smooth, linear progression. His seminal work, "The Structure of Scientific Revolutions," presents the concept of 'normal science' disrupted by 'revolutionary science,' leading to a new paradigm when the existing one fails to explain emerging anomalies. This perspective not only shifted the philosophy of science but also provided a framework for understanding progress in various disciplines, including network analysis. In this context, the integration of Kuhn's theory with network analytical tools like the I-indicator, C-indicator, and ICCO ranking offers a novel lens to evaluate the evolution and potential of nodes within a network, metaphorically reflecting the progression of scientific paradigms.

Thomas Kuhn's philosophy of science is a testament to the complex, often turbulent nature of scientific progress. He delineated the process of scientific evolution into two distinct phases:

Normal Science: A period where scientific activity operates under the confines of an existing paradigm, which dictates the standards for legitimate work within the community.

Revolutionary Science: A transformative phase that occurs when the prevailing paradigm can no longer reconcile anomalies, leading to a paradigm shift and the adoption of a new framework that redefines the scientific truth.

The transition from normal to revolutionary science in Kuhn's model is not just a shift in methodologies or techniques but a profound change in the worldview of the scientific community. This transformative journey mirrors the dynamics observed in network structures, where nodes (akin to scientific theories or practices) rise in influence or become obsolete as the overall network (the scientific community) evolves.

The network metrics of I-indicator, C-indicator, and ICCO ranking can be analogously seen through the Kuhnian lens:

I-indicator: This metric represents the potential for growth within a node, derived from its betweenness centrality relative to its node degree. It mirrors Kuhn's notion of scientific anomalies; as a node gains centrality without a proportional increase in connections, it may signify a potential for significant change—a harbinger of a 'revolution' within the network.

C-indicator: The clustering coefficient reflects the tendency of nodes to create tightly-knit groups. Similar to Kuhn's 'normal science,' nodes within a high C-indicator operate within a well-established network paradigm, reinforcing existing connections rather than fostering new ones.

ICCO Ranking: This composite metric assesses a node's significance based on the I-indicator, C-indicator, and the frequency of associated keywords. It is akin to the comprehensive approach needed to evaluate the viability of a scientific paradigm, taking into account not just the core structure (I-indicator and C-indicator) but also the node's contextual relevance (keyword frequency).

The application of Kuhn's theory to network analysis offers a multidimensional understanding of nodes and their roles within a network. Just as a scientific paradigm guides the methodology and focus of research, the I-indicator can determine how resources within a network should be allocated to maximize potential. Similarly, just as anomalies within a paradigm can lead to a scientific revolution, nodes with a high I-indicator but low degree may signal emerging influencers or shifts in the network.

The C-indicator, in Kuhnian terms, ensures that during periods of 'normal science,' the network maintains robustness and interconnectedness, much like the scientific community adheres to a paradigm to further develop its theories. Yet, it is the potential for interconnected nodes to form new patterns that can precipitate a paradigm shift, indicative of a network's evolution.

Lastly, the ICCO ranking system provides a holistic assessment similar to how Kuhn's model necessitates considering both the historical progression and the sociological aspects of science to fully understand the development of scientific knowledge. This comprehensive ranking enables the identification of nodes that could lead to the next paradigm shift within the network.

4.9. Application of machine learning algorithms

Subsequent to conducting a rigorous node analysis, my investigative efforts bore fruit in the form of identifying the most pivotal nodes within the network. These pivotal nodes served as the linchpin for my subsequent strategic endeavors, where I undertook a multifaceted approach to both harness the insights gleaned from these nodes and further explore the myriad facets of network features. The ensuing discussion provides a comprehensive overview of my investigative strategies:

First and foremost, my primary investigative avenue entailed the application of sophisticated deep learning models to the network data. my toolkit comprised a spectrum of state-of-the-art models, including Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), and Deep Graph Neural Networks, as well as innovative hybrid models that amalgamated the strengths of these individual approaches. This amalgamation allowed us to comprehensively scrutinize the network's complexities, unearthing hidden patterns and nuanced relationships.

My overarching objectives for deploying these cutting-edge models were twofold. Firstly, I sought to distill the rich information embedded within the network data to generate concise summaries of the textual content, thereby offering a concise yet informative encapsulation of its salient aspects. This summary generation aimed to facilitate efficient comprehension and decision-making within the domain of network analysis.

Secondly, I aspired to leverage the insights gleaned from the top nodes, which had been identified through the node analysis, to generate novel sentences. This innovative approach aimed to augment the utility of the network data by providing an avenue for the automatic generation of new content based on the underlying structure and significance of these nodes. The goal here was to foster creativity and expand the informational scope of the network analysis, potentially uncovering novel narratives or insights not readily apparent through conventional means.

Long Short-Term Memory networks (LSTMs), celebrated for their proficiency in handling sequential data, have found a compelling niche within the sphere of graph and network analysis. LSTMs offer a potent tool to model the dynamic facets intrinsic to networks, rendering them highly suitable for the analysis of time-series data and the exploration of dynamic graph structures. Their capability to capture temporal dependencies allows for a deeper comprehension of evolving network behaviors and patterns over time.

Temporal dependencies are pervasive in network data, often originating from evolving connections and interactions. LSTMs excel in capturing these dependencies, facilitating the modeling of dynamic aspects such as node interactions, link formation, and information propagation. This paper delves into how LSTMs can encode and decode temporal information within networks, offering insights into their application in various domains. Notably, LSTMs prove instrumental in analyzing time-series data, where networks undergo transformations over discrete time intervals, as well as in dynamic graph analysis, where the network structure evolves continuously.

However, applying LSTMs to network data presents challenges. Data preprocessing requires meticulous handling to align temporal data appropriately and extract relevant features. Additionally, network representation techniques, including node embeddings and graph encodings, are pivotal in facilitating LSTM-based analysis. Despite these challenges, the integration of LSTMs within the realm of network analysis holds immense potential, promising to advance my understanding of dynamic network behaviors across a spectrum of domains.

Subsequently, I embarked on a journey into the realm of Deep Graph Neural Network (DGNN) models, a cutting-edge field in the domain of machine learning and graph analysis. my research ventured beyond the conventional by exploring novel avenues for model enhancement. In particular, I sought to fuse DGNN with advanced composite models, such as the Deep Convolutional Graph Neural Network (DCGNN).

Deep Graph Neural Networks represent a pivotal paradigm shift in my approach to solving complex graph-related problems. These models possess the unique capability to capture intricate relationships and dependencies within large-scale graph data. However, recognizing the potential for further improvement and innovation, I embarked on the task of synergizing DGNN with composite models, like DCGNN, which incorporate convolutional neural network principles into the graph analysis domain.

The Deep Convolutional Graph Neural Network, or DCGNN, stands as a testament to my commitment to pushing the boundaries of what is achievable in graph-based machine learning. DCGNN's utilization of convolutional operations enables it to extract rich spatial features from graph data, allowing for more nuanced and accurate modeling. This innovation brings forth a host of advantages that set it apart from traditional DGNN models.

One of the standout advantages of DCGNN is its ability to capture local structural information efficiently, making it particularly adept at handling graphs with intricate topologies. Furthermore, its integration of convolutional layers allows for hierarchical feature extraction, enabling the model to discern patterns and relationships at multiple scales within the graph. This multi-scale analysis empowers DCGNN to excel in tasks such as node classification, link prediction, and graph classification.

As previously mentioned, the overarching objective driving the implementation of these deep learning models centers on their capacity to generate concise summaries and construct novel sentences by leveraging the most salient nodes within the dataset. The models I meticulously designed have indeed yielded remarkable results and have demonstrated a notable degree of alignment with the underlying dataset. However, it is imperative to recognize that the realm of natural language processing (NLP) tasks encompasses a spectrum of intricate challenges that extend beyond the realm of conventional accuracy and traditional metrics. These challenges necessitate the fulfillment of additional criteria to ensure the generation of high-quality and practically useful outputs.

Within the purview of NLP, achieving proficiency transcends mere numerical assessments, extending into the realm of human interpretability and readability. In essence, a model's merit cannot be solely gauged by its performance metrics but is intrinsically linked to its ability to produce outputs that are not just computationally correct but are also comprehensible and intelligible to human users. In practical terms, this implies that while my model may exhibit exceptional prowess when measured against established quantitative metrics, its practical utility comes under scrutiny when it encounters real-world scenarios featuring novel or previously unseen input data.

Henceforth, a pivotal imperative arises: in addition to demonstrating commendable results based on quantitative metrics, my NLP model must possess the crucial ability to produce output that not only meets high-quality standards but is also readily interpretable by a diverse user base. This user base encompasses experts with deep domain knowledge as well as non-experts with more generalized familiarity. The crux of the matter is to ensure that the model's output transcends mere data-driven proficiency and aligns with the pragmatic requirements of users from various backgrounds and levels of expertise.

My deployed models yielded commendable results in terms of quantitative metrics. However, a critical examination of their performance when confronted with previously unseen inputs reveals a significant limitation related to the human interpretability of their outputs, particularly in the specific context of generating novel sentences. While these models excel in summarizing text, demonstrating their proficiency by autonomously selecting and presenting the most salient sentences and words, their effectiveness in generating entirely new sentences presents a distinctive challenge.

In the realm of text summarization, my models exhibit exceptional prowess by obviating the need for explicit input selection. Instead, they automatically identify and extract the most relevant sentences and words, consistently producing high-quality summaries. This capability is a testament to the models' capacity to independently discern crucial content within a given text, leading to favorable outcomes.

However, the dynamics change when I shift my focus to the generation of novel sentences. In this scenario, the models are constrained to construct sentences based on predetermined top nodes, which emerge from intricate network analyses. This constraint introduces a notable complexity, as the models must adhere to the network-driven criteria when generating sentences. Consequently, this constraint may occasionally lead to output sentences that, while aligning with the network's specifications, may not consistently yield outputs that are coherent or readily comprehensible from a human perspective.

In addition to the bespoke deep learning models previously introduced and applied in my research, I sought to harness the capabilities of pretrained Large Language Models (LLMs) as part of my methodology. This strategic approach offers substantial advantages over the custom models I developed in-house. Notably, pretrained LLMs come equipped with the significant benefit of prior training, eliminating the need for extensive computational resources during model development and enabling us to leverage their vast knowledge and linguistic proficiency.

My endeavor to incorporate pretrained LLMs into my study encompassed two distinct approaches. Firstly, I opted to employ these pretrained models directly for my tasks, which included tasks such as text summarization and the generation of new sentences. In this approach, I harnessed the inherent capabilities of these pretrained LLMs without further fine-tuning or customization. This "out-of-the-box" utilization capitalizes on the models' extensive training on diverse text corpora, allowing them to exhibit impressive linguistic prowess.

Secondly, I embarked on a more tailored path by subjecting the pretrained LLMs to additional training based on my custom datasets. These datasets comprised a blend of academic and non-academic content, designed to align more closely with the specific objectives of my study. This approach allowed us to fine-tune the pretrained models to cater to the intricacies and nuances inherent in my research domain, enhancing their suitability for my tasks.

In order to harness the capabilities of Large Language Models (LLMs), I integrated the OpenAI module into my Python-based workflow, providing us with a versatile toolset for invoking and leveraging these LLMs. This module encompasses a variety of LLM variants, affording us the flexibility to select and employ models best suited to my specific research objectives.

Notwithstanding the considerable utility of LLMs, it is imperative to acknowledge certain inherent limitations that warrant attention. These drawbacks are enumerated as follows:

Limited Availability of Latest Models: One noteworthy constraint pertains to the availability of the most recent LLM iterations. For instance, my ability to utilize state-of-the-art models like GPT-4 or even GPT-3.5 may be constrained. Consequently, I may find myself compelled to work with older versions of LLMs. While these older models remain formidable, their computational prowess and capabilities may not measure up to the cutting-edge iterations.

Domain Discrepancies: Another key consideration revolves around the alignment of LLMs with the thematic scope of the task at hand. LLMs are trained on vast and diverse corpora, and their efficacy can be compromised when confronted with tasks that diverge significantly from the topics they have been exposed to during training. This issue becomes particularly pronounced when employing older LLM versions, further diminishing their reliability and efficiency in such contexts.

Custom Dataset Training: Should I opt to fine-tune LLMs using custom datasets, this process necessitates substantial computational resources. However, the effectiveness of this fine-tuning endeavor is contingent upon the size and representativeness of the dataset. Inadequate dataset size can significantly limit the benefits derived from custom training.

Despite these constraints, LLMs, by virtue of their computational prowess and linguistic sophistication, have proven to be instrumental in my research efforts. Comparatively, they have outperformed my custom models, underscoring their considerable utility. However, it is crucial to emphasize that even these powerful LLMs, while delivering noteworthy results, have not consistently yielded outcomes of a magnitude that could be described as significant or extraordinary within the context of my research objectives.

At the heart of my approach lies the formidable GPT 3.5 model, renowned for its prowess in natural language generation. With its advanced architecture and deep understanding of linguistic nuances, this model became my trusted ally in generating advice that resonated with project managers' needs. By capitalizing on the vast knowledge encapsulated within the top nodes, I sought to offer not just generic recommendations but insightful guidance derived from the unique dynamics of each project.

The integration of the GPT assistant and pretrained LLMs represents a paradigm shift in project management practices. It allows project managers to benefit from automated assistance that complements their expertise and augments their decision-making processes. The generated advice serves as a compass, pointing project managers in the right direction as they navigate through complexities, anticipate risks, allocate resources, and foster collaboration within their teams.

Moreover, the versatility of language models and the adaptability of deep learning algorithms open up exciting possibilities for future advancements in this field. As technology continues to evolve, project managers can anticipate even more accurate, nuanced, and context-sensitive advice that caters to the unique requirements of each project scenario.

In my relentless pursuit of advancing the advice generation process for project managers, I embarked on an ambitious and comprehensive endeavor that encompassed a multi-faceted approach. my methodology involved not only the integration of custom datasets but also the exploration and utilization of various techniques to maximize the potential of pretrained large language models (LLMs). By combining these elements, I aimed to deliver a holistic and highly effective solution for generating tailored and contextually relevant advice in the field of project management.

The inclusion of custom datasets played a pivotal role in enriching the knowledge base from which my advice generation system drew insights. I recognized the importance of capturing a diverse range of information sources to ensure a deep understanding of the intricacies and complexities of project management. To achieve this, I curated a comprehensive collection of data that spanned academic research papers, industry reports, best practices, case studies, and real-world project experiences. This multidimensional approach allowed us to tap into a vast reservoir of knowledge and expertise, ensuring that the generated advice would be well-rounded and applicable to a wide array of project scenarios.

An integral part of my methodology was the training of custom datasets using pretrained LLMs. These language models, such as GPT 3.5, had undergone extensive pre-training on massive amounts of text data, enabling them to master the subtleties of language and understand complex linguistic patterns. By fine-tuning these models with my custom datasets, I harnessed their powerful capabilities to generate advice that was not only relevant but also aligned with the unique requirements of project management.

The advantage of leveraging pretrained LLMs lay in their ability to adapt to novel datasets and generate high-quality output. Their capacity to comprehend context, grammar, and semantics allowed them to produce coherent and readable sentences that closely resembled human-authored text. This inherent understanding of language structures ensured that the generated advice was not only insightful but also easily understandable for project managers seeking guidance. The robustness and comprehensiveness of the pretrained models, derived from training on vast corpora of diverse text, provided a strong foundation for generating well-structured and contextually grounded advice.

However, I acknowledge that challenges exist when working with custom datasets and pretrained models. Limited availability of domain-specific data and resource constraints can result in smaller dataset sizes compared to those used to train models like GPT 3.5. This discrepancy may impact the models' ability to capture rare or specific project management scenarios, potentially leading to a slight decrease in overall performance and accuracy of the generated advice. Moreover, refining and enhancing the models with scarce data can pose a unique set of difficulties, as the models might find it challenging to extrapolate beyond the distinct patterns observed in the training data.

Despite these challenges, I firmly believe in the value and potential of combining custom datasets with pretrained LLMs. Even with a constrained dataset, these models retain the capacity to provide valuable insights and guidance to project managers, serving as powerful decision support tools. By leveraging post-processing techniques, such as fine-tuning and optimization, I can further enhance the quality, relevance, and specificity of the generated advice. Furthermore, ongoing advancements in computational resources, data availability, and model architectures will contribute to narrowing the gap between pretrained models and state-of-the-art approaches like GPT 3, pushing the boundaries of what can be achieved in project management advice generation.

In the exciting world of natural language processing, a range of pretrained Language Models (LLMs) are available to generate sentences, each offering its own unique capabilities. Let's delve into three notable examples: GPT-2, T5 developed by Google, and BART developed by Facebook.

GPT-2, renowned for its impressive performance, has gained significant recognition in the field. Its strength lies in generating coherent and contextually relevant sentences, making it especially valuable for tasks like text completion and creative writing. However, due to its large size and complexity, GPT-2 may be computationally expensive and requires substantial computational resources.

T5, developed by Google, stands out with its versatility and flexibility. It encompasses a wide range of tasks, including text summarization, translation, question answering, and more. T5 offers the advantage of fine-tuning, allowing users to adapt the model to specific domains or tasks. However, this flexibility may come at the cost of increased training time and potentially reduced performance in certain specialized tasks.

BART, an acronym for Bidirectional and Auto-Regressive Transformers, is a LLM introduced by Facebook. One distinctive feature of BART is its ability to handle both generative and reductive tasks effectively. With its denoising autoencoder architecture, BART excels at tasks like text summarization and text generation from prompts. Nevertheless, compared to other models, BART might require more extensive fine-tuning to achieve optimal performance in different applications.

It's important to note that while these LLMs offer great potential, they also have their limitations. Generating coherent and contextually accurate sentences remains a challenge, particularly when faced with ambiguous or complex inputs. Furthermore, the ethical considerations surrounding bias and fairness in language generation algorithms are ongoing concerns that need careful attention.

In the following section, I have presented a selection of sentences generated by both GPT 3 and trained models, showcasing their respective capabilities and differences in output:

GPT 3:

Generated text by academic (Title Unigram section) dataset:

Approach analysis is a project management technique used to analyze the various approaches that may be used to complete a project. It is used to identify the most efficient and effective approach for completing the project. Approach analysis can help project managers identify potential risks, determine the best resources to use, and identify any potential problems that may arise. It can also help to identify the most cost-effective approach and the most appropriate timeline for the project.

Agile project management is an iterative, adaptive approach to project management that is based on delivering requirements in small increments. Agile project management focuses on collaboration, communication, and continuous improvement, and emphasizes the importance of customer feedback in the development process. Projects in the field of project management can range from large-scale construction projects to small-scale software development projects. Some examples of project management projects include:

- Developing a project charter to define the scope and objectives of a project.
- Creating a project plan to identify tasks, resources, and timelines.
- Managing stakeholder expectations and communications.
- Developing a risk management plan to identify and mitigate risks.

Model development analysis is a useful tool in the field of project management. It is a process for analyzing a project or system and developing a model of the system or project

Agile study construction in the field of project management is the process of designing and carrying out research studies that are based on the principles of Agile project management. This approach to research involves a focus on collaboration between stakeholders, rapid prototyping, and iterative development of research questions and methods. It also emphasizes the importance of user feedback and continuous improvement. Agile study construction can be used to identify and address user needs and to develop a better understanding of project management processes and practices.

Project management is the process of initiating, planning, executing, controlling, and closing the work of a team to achieve specific goals and meet specific success criteria. The primary challenge of project management is to achieve all of the project goals within the given constraints. This includes the constraints of time, cost, quality, scope, resources, and risk.

In the field of project management, software study construction can help with the organization and management of projects.

Agile project management is a project management approach that focuses on delivering value through rapid, iterative, and incremental work cycles. It is based on the principles of agile software development and is designed to help teams respond quickly to changing customer needs and market conditions. Agile project management emphasizes collaboration, continuous improvement, and flexibility. It utilizes short development cycles, called sprints, to ensure that projects remain on track and that customer feedback is incorporated quickly. Agile project management can help teams deliver products

There are a number of software systems available in the field of project management. Popular software systems include Microsoft Project, Clarizen, Wrike, Basecamp, Asana, and Trello. Each of these systems provides features such as task management, project planning, collaboration tools, and resource management. Depending on the specific needs of your project, one of these systems may be the best fit.

The case-based approach in project management is a method of problem solving that involves studying and analyzing past cases in order to identify the best strategies and solutions for similar situations. This approach is especially useful for complex projects, as it allows project managers to draw on the experiences of other projects to help inform their decisions. By studying the successes and failures of similar projects, project managers can gain insights into how to approach their own projects and avoid potential pitfalls.

Risk control in project management is an interdisciplinary field that merges project management, computer science, and business. It encompasses the utilization of computer systems and software to detect and manage potential risks linked to a project. This process includes the detection of possible risks, the assessment of their influence on the project, and the formulation of strategies to lessen their impact. It also involves the implementation of systems and processes to monitor and control risks throughout the project lifecycle.

Project risk management is a process that involves the identification, assessment, and mitigation of risks associated with a project. It is an essential part of project management and involves the use of various tools and techniques to reduce the likelihood of project failure. In the field of project management, computer systems can be used to help manage risks. These systems can provide automated risk assessment and analysis, as well as automated risk mitigation plans. They can also provide real-time updates on project progress, allowing project managers to Interdisciplinary and multidisciplinary approaches to project risk management involve the coordination of multiple disciplines and stakeholders to identify, analyze, and manage risks. This includes the integration of expertise from different fields such as engineering, finance, economics, and law. The goal is to develop a comprehensive risk management strategy that considers the perspectives of all stakeholders and disciplines. This approach can help to ensure that risks are identified and managed effectively, and that projects are successful.

Generated text by academic (Title Bigram section) dataset:

A web-based construction industry information system in the field of project management can be used to track and manage projects, provide access to project documents and plans, and monitor project progress. It can also be used to store and share project data, manage project budgets, and track project costs. Additionally, the system can be used to communicate with stakeholders, track project milestones, and generate reports. The system can be used to collaborate with other project stakeholders, such as architects, engineers, and contractors.

The critical success maturity level of a new product in the field of project management is highly dependent on the product's ability to meet the needs of its users. A successful product should be able to provide a comprehensive set of project management tools, such as task management, resource management, budgeting, and reporting. Additionally, the product should be able to integrate with existing systems and provide a high level of scalability.

A web-based learning construction industry information system in the field of project management could provide a comprehensive database of industry-specific information, including resources for project management, best practices, and industry news. It could also provide interactive tools for project management, such as project planning, scheduling, budgeting, and reporting. The system could also include a library of relevant articles, case studies, and research materials. Finally, the system could offer online courses and certification programs to help users improve their project management.

This includes keeping stakeholders informed of progress, changes, and updates, as well as ensuring that team members are aware of their responsibilities and tasks. Establishing clear goals and objectives for the project is essential to its success. By setting measurable objectives and timelines, project managers can ensure that all stakeholders are.

Generated text by academic (Keywords Bigram section) dataset:

Project management is a multidisciplinary and interdisciplinary field that seeks to bring together the various disciplines necessary for the successful completion of a project. This includes engineering, science, sustainability, economics, and other disciplines. Sustainable project management involves the integration of sustainability principles into the project management process. This includes the use of renewable energy sources, the reduction of waste, and the use of sustainable materials and technologies.

Sustainable science engineering multidisciplinary energy fuel project management involves the application of project management principles and techniques to the development of sustainable energy sources and fuels. This includes the development of renewable energy sources such as solar, wind, and geothermal, as well as the development of alternative fuels such as biofuels, hydrogen, and synthetic fuels.

Project management theory is an interdisciplinary field that applies knowledge, skills, tools, and techniques to project activities in order to meet project requirements. It is an evolving field that is constantly evolving and adapting to new challenges. The main goal of project management theory is to ensure successful completion of projects in a timely and cost-effective manner.

Project management is a field that relies heavily on the use of electrical theory, methods, and science. This includes the use of mathematical models, algorithms, and computer simulations to develop project plans, schedules, and cost estimates. Electrical theory can be used to develop efficient and effective project plans and schedules. Methods such as linear programming and network analysis can be used to optimize resources and cost estimates. Science theory can be used to develop risk assessments and understand the behavior of complex systems.

Green sustainable engineering is an emerging multidisciplinary field of engineering that focuses on creating and managing projects that are both sustainable and environmentally friendly. It combines the principles of electrical engineering, project management, and sustainability to create innovative solutions to environmental challenges. This field of engineering is growing in popularity as more organizations and individuals become aware of the importance of sustainability. Green sustainable engineering focuses on creating projects that are not only efficient and cost-effective, but also have a positive impact on the environment. Examples of

Project management is a multidisciplinary field that combines engineering, operations research, and other disciplines to ensure successful project completion. It involves the application of knowledge, skills, tools, and techniques to project activities in order to meet project requirements. Project management involves the planning, organizing, directing, and controlling of resources and activities to achieve specific project objectives. It includes processes such as project initiation, planning, execution, control, and closeout.

Project management is an increasingly important field in sustainable science, as it is used to ensure that projects are completed efficiently, on time, and within budget. Sustainable science in project management involves the use of methods and tools that are designed to minimize environmental impact, reduce costs, and increase efficiency. This includes the use of green technology, life cycle assessment, and other techniques that aim to reduce waste and conserve resources. Additionally, project management tools such as LEED certification, carbon footprinting, and energy

Project management in the energy and fuel science field is an important area of research. It involves the development of strategies and methods to manage projects related to energy and fuel science. This includes understanding the dynamics of energy and fuel production, storage and use, as well as the economic, environmental and social impacts of these activities. Projects may include the development of new technologies, the implementation of new policies and regulations, or the development of efficient and sustainable energy and fuel systems.

Generated text by academic (Keywords Unigram section) dataset:

Project management engineering systems are designed to help manage projects by providing a system of tools and processes to help plan, manage, and control the project. These systems typically include project planning tools, project scheduling tools, resource management tools, risk management tools, and reporting tools.

Project management theory research can include topics such as the application of existing project management principles and theories, the development of new project management theories, and the evaluation of the effectiveness of project management practices. It can also include research into the impact of project management on organizational performance, the effectiveness of project management tools and techniques, and the role of project managers in successful project delivery.

Project economics is the application of economic analysis to the evaluation and management of projects. It is a subset of project management and involves the use of various methods to analyze the economic feasibility of a project. These methods include cost-benefit analysis, discounted cash flow analysis, risk analysis, and sensitivity analysis.

Project management is an area of economics that examines the application of scientific principles and techniques to the management of projects. Economic principles can be used to analyze the cost and benefit of different project management approaches, such as cost-benefit analysis, break-even analysis, and risk analysis.

Multidisciplinary system science in the field of project management is the application of a wide range of disciplines to the study of project management. This includes fields such as psychology, engineering, economics, sociology, anthropology, and others. This approach seeks to understand how various disciplines can be used to improve the management of projects, from planning to execution.

Interdisciplinary multidisciplinary engineering in the field of project management involves the use of a variety of engineering disciplines to develop project management solutions that are tailored to the specific needs of a project. This could include the use of civil engineering, mechanical engineering, electrical engineering, industrial engineering, and other disciplines to develop solutions for managing a project from start to finish.

Project management is an interdisciplinary field that draws on many different disciplines, such as economics, psychology, sociology, mathematics, and engineering. Research in project management has focused on topics such as risk management, cost estimation, scheduling, resource allocation, project control, and project success.

Automated Scheduling Tools: Automated scheduling tools can be used to create project plans, track progress, and maintain accountability.

Cloud Collaboration Tools: Cloud collaboration tools allow project teams to work together in real-time from different locations. This can help project managers keep everyone on the same page and ensure that tasks are completed quickly and efficiently.

4.10. Using machine learning algorithms as new metric

My methodology is grounded in the recognition that the frequency of keyword occurrence is instrumental in determining its relevance and, consequently, its importance for my ongoing research.

My preliminary step involves the identification of the top 100 most frequently occurring keywords. This specific subset of words warrants my highest level of attention and forms the cornerstone of my subsequent analysis, given their pronounced impact on the subject matter.

To maintain consistency and allow for accurate comparative analysis, I have adopted a quarterly temporal framework. Each quarter serves a specific purpose in my study. The initial quarter is referred to as the "base quarter," while the succeeding quarter is utilized for comparative analysis and is termed the "comparison quarter."

Labeling is conducted in the following manner, contingent on the movement of the keywords between quarters:

- **Label "1":** Assigned to keywords that were previously outside the top 100 in the base quarter but have ascended into this paramount group in the comparison quarter, indicating a surge in relevance.
- **Label "-1":** Designated for keywords that occupied a position within the top 100 in the base quarter but have declined in ranking, consequently falling out of this elite categorization in the comparison quarter, signifying a decrease in importance.
- **Label "0":** Attributed to two distinct groups of keywords. The first group encompasses keywords consistently positioned within the top 100 across both quarters, denoting stability in their relevance. The second group includes those that remain outside the top 100 in both quarters, reflecting a continued lack of sufficient prominence to be considered for further analysis.

This labeling strategy is integral to my analytical processes, serving as a conduit for the systematic and quantitative assessment of data trends. It assists us in tracking the dynamism inherent in keyword significance, further informing my research direction and focus.

I believe this rigorous methodology affords us the precision necessary in my study and aids in the nuanced understanding of my field of inquiry. I remain open to further optimizing this system based on insights and findings as my research progresses. In the following section comprehensively described about this methodology.

In this research scenario, I undertook a comprehensive exploration of the results emanating from my network analysis endeavors. As mentioned earlier, my study was structured around a temporal framework consisting of ten discrete periods, each spanning a three-month interval from 2021 to 2023. Within this temporal landscape, a central objective crystallized: to decipher the intricacies responsible for the fluctuations in rankings experienced by specific nodes over these ten periods. my overarching aim was to unravel the underlying causal factors that contributed to either the ascent or descent in the rank of individual nodes over time. In essence, my research sought to identify the analytical methodologies capable of effectively forecasting the future trajectory of a given node.

This particular phase of my research was exclusively conducted within the non-academic dataset, marking a deliberate choice that bears further elucidation. my decision to focus specifically on the title section of the non-academic dataset stemmed from several considerations. First and foremost, the title section afforded a more manageable dataset size compared to the voluminous body of textual data. This pragmatic choice was driven by the need to streamline my analytical efforts and circumvent potential computational bottlenecks associated with working on the entire body of data.

Moreover, the non-academic dataset stood out for its propensity to provide us with information that held a higher degree of utility and relevance for my analytical pursuits. The diverse nature of the content within this dataset presented us with valuable insights that were instrumental in deciphering the underlying patterns governing the fluctuations in node rankings.

To arrive at these insights, I meticulously devised a robust and systematic methodology. For each of the ten periods, I implemented a stringent filtering process aimed at identifying and isolating the top 100 nodes based on their degree centrality. This approach effectively pinpointed the nodes, or n-grams, that exhibited the highest degree of connectivity within the dataset, signaling their frequent occurrence and influence. As a result, these nodes were accorded prominent positions in the ranking hierarchy, forming the focal points of my analysis.

Within one of the methodological scenarios I explored, my focus was primarily directed towards the examination of the 100 top nodes within each distinct period. In this endeavor, my objective was to extract nodes that exhibited a consistent presence across multiple periods while demonstrating discernible ascending or descending trends in their rankings. Regrettably, the cohort of nodes displaying these specific characteristics proved to be relatively limited in size. Consequently, the insights derived from this approach were insufficient to yield a comprehensive understanding of the underlying analytical patterns.

Subsequently, I embarked on another methodological trajectory to enhance my understanding of node dynamics over time. In this alternative approach, I designated nodes from the periods in 2021 (comprising four distinct periods) as input data and juxtaposed them with nodes from the 2023 periods (encompassing two periods), treated as output data. This approach aimed to discern and quantify the relationships that existed between nodes during these time intervals. Specifically, my analysis delved into nodes originating from the 2021 periods that exhibited distinctive trends, such as ascending or descending trajectories. It is imperative to note that this phase of the study extended its purview beyond the confines of the top 100 nodes, encompassing the entirety of available data.

Furthermore, I explored the notion of nodes that experienced a transition in their ranking status, which was of particular interest. These nodes, for instance, held a position beyond the top 100 rankings during the 2021 periods but subsequently ascended in prominence, securing a place among the top 100 nodes in the 2023 periods. This phenomenon prompted a detailed investigation into the predictive capabilities of various network analysis metrics.

Through rigorous analysis, I unearthed valuable insights. Specifically, for nodes that initially held rankings beyond the top 100 but exhibited an ascent to the top 100 in the 2023 periods, certain network analysis metrics such as load centrality, hub centrality, authority centrality, and ICCO ranking displayed the capacity to predict the trend trajectory of these nodes. However, it is noteworthy that in the converse scenario—nodes descending in rank—these analytical metrics proved less effective in predicting the trend.

Across the spectrum of methodological scenarios I meticulously explored, my central mission was to unearth analytical methodologies capable of reliably predicting the trajectories of node rankings, particularly in terms of nodes displaying both ascending and descending patterns. Nonetheless, each of these scenarios posed its distinct set of intricacies and challenges, exemplifying the intricate nature of this analytical pursuit.

In my initial scenario, my endeavor was to encompass the entirety of my dataset, spanning from the initial period in 2021 to the culmination of the study in 2023. Within this expansive temporal span, I aimed to identify nodes exhibiting clear ascending or descending trends. However, this undertaking encountered a formidable challenge: the dearth of nodes meeting these precise criteria. Regrettably, the limited number of nodes that displayed the desired trends hindered my ability to glean comprehensive insights, and the resulting dataset did not yield a holistic view of the underlying dynamics.

Subsequently, in the scenario where I structured my analysis around nodes from the 2021 periods as input and nodes from the 2023 periods as output, a nuanced distinction emerged in the efficacy of my selected analytical metrics. Notably, in the context of ascending trends, the analytical metrics exhibited a higher degree of predictive accuracy, facilitating the discernment of node trajectories. However, this predictive capability exhibited a degree of faltering when applied to nodes displaying descending trends. This variance underscored the intricate dynamics of network properties across distinct temporal periods, encompassing fluctuations in node sizes and the underlying network structures.

In light of the multifaceted challenges encountered in my analytical pursuits, it becomes evident that an exclusive reliance on individual analytical metrics may not yield the optimal results. Given the dynamic nature of network properties over time, a more promising avenue lies in the exploration of potential interactions between various analytical approaches. By adopting this holistic perspective, I can harness the synergistic potential of different metrics, thereby enhancing my capacity to predict node trends effectively.

Regrettably, my exploration of the previously discussed scenarios did not yield the desired outcomes in terms of providing reliable indicators for my research task. Faced with this challenge, I opted for a strategic pivot, redefining my approach by framing the task as a classification problem. This transition allowed us to leverage the formidable capabilities of both machine learning and deep learning algorithms, presenting a fresh perspective on the inherent complexities of my research endeavor.

A salient distinction between this revised scenario and its predecessors is worth emphasizing. In my earlier scenarios, my focus was predominantly univariate, centering on the assessment of individual analytical metrics in isolation. In stark contrast, this newly adopted classification approach adopts a more comprehensive outlook by incorporating the entire spectrum of available analytical metrics as input features for my designed models. This strategic shift signifies a profound departure from my prior strategies, as it enables us to harness the collective insights emanating from a diverse range of analytical perspectives.

Furthermore, it is paramount to underscore that within the context of the classification task, my exclusive attention was directed towards the domain of network analysis. Herein lies another departure from my previous approaches. I consciously chose to exclude certain non-compatible analyses that fell outside the purview of Networkx, such as the I-indicator, ICCO, ICCO ranking, and analogous metrics. This omission was underpinned by a fundamental recognition that these specific analyses could be synthesized as composite measures, thereby eliminating the need to incorporate them as distinct metrics. In essence, my strategy involved streamlining my analytical approach by focusing on the essential network analysis metrics that promised to provide the most salient insights.

In the context of my classification task, I adopted a structured approach wherein the various analyses served as feature variables, while my target variable consisted of three distinct classes, namely, -1, 0, and 1. The determination of these classes was rooted in my criterion of the top 100 nodes, which played a pivotal role in shaping my classification framework. These classes were derived through a meticulous process of comparing node rankings between consecutive pairs of periods within my temporal dataset.

My approach involved evaluating individual nodes on a one-word-at-a-time basis, tracking their presence or absence within the top 100 rankings across successive time periods. To elucidate this process, consider a node within the context of two consecutive periods. If a node occupied a position within the top 100 rankings during the first period but failed to maintain that status in the subsequent period, it was assigned a label of -1. Conversely, if a node was not within the top 100 rankings in the first period but ascended to the top 100 in the second period, it was labeled as 1.

In cases where a node maintained a consistent ranking status across both periods, meaning it remained within the top 100 in both or outside the top 100 in both, it was assigned the label 0. This classification served to capture nodes that exhibited stability in their rankings, with no significant ascent or descent in status between the two time periods.

This approach allowed us to systematically categorize nodes based on their ranking dynamics, thereby facilitating the creation of a meaningful target variable for my classification task. By assigning numerical labels to nodes in this manner, I established a framework to train and evaluate machine learning and deep learning models that could subsequently predict and classify the temporal behavior of nodes within the network.

To illustrate the labeling process, we present examples of how nodes were classified based on ranking changes between consecutive periods. The classification criterion focused on the top 100 nodes. We tracked nodes' presence within the top 100 across successive periods, assigning labels based on their ranking dynamics, which facilitated the creation of a meaningful target variable for classification.

Nodes were labeled as follows:
Label -1: A node in the top 100 in the first period but not in the second.
Label 1: A node outside the top 100 in the first period but in the top 100 in the second.
Label 0: A node consistently in or out of the top 100 across both periods.

This systematic approach enabled us to train and evaluate models predicting nodes' temporal behavior. See below the Example Nodes:

“Free” Node:

As Table. 1 shows the node "free" remained within the top 100 rankings, indicating sustained relevance, likely due to ongoing interest in concepts like liberty or free resources.

Table 1. “Free” Node Ranking in consecutive quarters (own research)

1st Quarter of 2021 Rank	2nd Quarter of 2021 Rank	Label
6	13	0

“Policy” Node:

As Table. 2 shows , the node "policy" dropped significantly, reflecting a decreased focus on policy-related discussions, possibly due to shifting priorities or emerging new issues.

Table 2. “Policy” Node Ranking in consecutive quarters (own research)

1st Quarter of 2021 Rank	2nd Quarter of 2021 Rank	Label
54	1325	-1

“Engineering” Node:

As Table. 3 shows, the node "engineering" rose dramatically, suggesting increased attention to technological advancements or infrastructure developments.

Table 3. “Engineering” Node Ranking in consecutive quarters (own research)

1st Quarter of 2021 Rank	2nd Quarter of 2021 Rank	Label
1806	25	1

“Background” Node:

As Table. 4 shows, the node "background" fell out of the top 100, indicating a shift away from foundational discussions to other emerging themes.

Table 4. “Background” Node Ranking in consecutive quarters (own research)

3rd Quarter of 2022 Rank	4th Quarter of 2022 Rank	Label
10	Doesn’t exist	-1

“Productivity” Node:

As Table. 5 shows, the node " Productivity" rose dramatically, suggesting increased attention to technological advancements or infrastructure developments.

Table 5. “Productivity” Node Ranking in consecutive quarters (own research)

3rd Quarter of 2022 Rank	4th Quarter of 2022 Rank	Label
2352	12	1

“Include” Node:

As Table. 6 shows, the node "include" remained outside the top 100, reflecting steady but secondary importance in discussions about inclusivity.

Table 6. “Include” Node Ranking in consecutive quarters (own research)

3rd Quarter of 2022 Rank	4th Quarter of 2022 Rank	Label
269	**1451**	**0**

These examples illustrate the dynamic nature of the dataset, showing how various topics evolved in importance over time. To enhance comprehension of the labeling procedure, we provide the pseudocode in Fig. 21.

```
Labelling Procedure

TOP_RANK_THRESHOLD = 100
def determine_label(rank1, rank2):
    if rank1 ≤ TOP_RANK_THRESHOLD and rank2 ≤ TOP_RANK_THRESHOLD:
        return 0  # Stayed in top 100
    if rank1 > TOP_RANK_THRESHOLD and rank2 > TOP_RANK_THRESHOLD:
        return 0  # Stayed outside top 100
    return -1 if rank1 ≤ TOP_RANK_THRESHOLD else 1  # Dropped out or Entered top 100
for node1 in two_consecutive_quarter_data:
    name = node1["name"]
    rank1 = node1["rank"]
    node2 = find_node_by_name(quarter2_data, name)
    rank2 = node2["rank"] if node2 else TOP_RANK_THRESHOLD + 1
    label = determine_label(rank1, rank2)
```

Figure 21. Labelling procedure pseudo code (own research)

It is imperative to underscore that the overall dataset size, encompassing the entirety of data collected across all periods, amounts to approximately five thousand data points. When I embarked on the process of labeling the nodes within this dataset, a noteworthy observation emerged: the vast majority of nodes received a label of 0. This outcome can be attributed to the inherent constraints of my classification scheme, which is intricately linked to the limited number of nodes that attain top rankings within each individual period.

Indeed, within each period, I consistently identified only a select subset of nodes—specifically, the top 100 nodes—as candidates for inclusion in my classification scheme. Under the most optimistic circumstances, this would lead to the creation of a mere 100 classes with a label of 1, denoting nodes that ascended from non-top-100 status to the top 100. Similarly, there would be another 100 classes with a label of -1, signifying nodes that descended from top-100 status to non-top-100 status in the subsequent period. The remaining nodes, constituting the majority, would inevitably be assigned the label 0.

This distribution reflects the inherent rarity of nodes that exhibit significant changes in their ranking status between consecutive periods. The limited number of top positions, coupled with the intricate dynamics of network behavior, results in a predominantly stable node population, thus giving rise to the preponderance of nodes assigned the label 0.

In essence, this classification scheme encapsulates the fundamental nature of my data, characterized by its sparsity in terms of nodes experiencing substantial shifts in their rankings. Consequently, the task of predicting and classifying nodes exhibiting pronounced temporal dynamics remains a challenging endeavor within this context.

Henceforth, I are confronted with an inherently imbalanced class classification problem, a circumstance that can exert a profound impact on the performance of my predictive models and subsequent results. This class imbalance presents a formidable challenge, necessitating thoughtful mitigation strategies to ensure the robustness and fairness of my classification framework.

Upon the completion of my node labeling process, I conducted a meticulous assessment of the size distribution within each class, particularly focusing on the nodes designated with labels 1 and -1. It became evident that these two classes, representing nodes that exhibited significant changes in their ranking statuses, were inherently smaller in size due to the relative rarity of such dynamics within my dataset.

In light of this class imbalance, it was imperative to adopt measures aimed at reducing the size of the predominant 0-labeled class, thus rectifying the skewed distribution. my approach to address this issue involved a careful consideration of class size equalization. In essence, I aimed to bring the class sizes of 1-labeled and -1-labeled nodes in line with that of the 0-labeled class. To achieve this balance, I strategically selected a random subset of nodes from the 0-labeled class, with the size of this subset mirroring that of the smaller 1-labeled and -1-labeled classes.

This meticulous class size adjustment procedure was implemented with the intention of fostering a more equitable and representative dataset for my classification task. By leveling the playing field in terms of class sizes, I sought to mitigate the undue influence of the majority class while affording my models a fairer opportunity to recognize and predict the nuanced patterns characterizing nodes undergoing significant ranking changes.

I systematically applied machine learning (ML) and deep learning (DL) models to both datasets, which encompassed instances with both imbalanced and balanced class distributions. The outcomes derived from these modeling efforts are meticulously detailed in Figure 4, thereby affording a comprehensive overview of the model performance across these distinct scenarios.

It is noteworthy that my expectations were substantiated by the outcomes of these analyses. By leveraging the comprehensive set of analytical features comprising all available analyses, my ML and DL algorithms demonstrated remarkable predictive prowess. This aptly resulted in a series of figures that elucidate the performance metrics and model efficacy across the various configurations.

The ensuing figures vividly encapsulate the fruits of my analytical labor, serving as visual representations of the models' performance across an array of pertinent metrics. These visual aids not only provide a succinct summary of my findings but also facilitate a deeper understanding of the nuanced trends and patterns observed within my results.

In the scope of my research, I undertook the utilization of a diverse array of computational algorithms, including XGboost, random forest, the multi-layer perceptron (MLP) neural network, and the Support Vector Machine (SVM). my primary objective was to develop a model that balances simplicity with robust predictive capabilities, thereby facilitating the achievement of my desired research outcomes.

Expanding upon this, my rationale for employing this ensemble of algorithms stemmed from the recognition that different machine learning approaches possess distinct strengths and weaknesses. XGboost and random forest, for instance, excel at handling complex, non-linear relationships within the data, while the MLP neural network exhibits remarkable adaptability for modeling intricate patterns. On the other hand, the SVM is particularly well-suited for classification tasks with high-dimensional feature spaces.

As previously indicated, my research endeavors involved tackling a challenging issue associated with imbalanced class distributions, a factor that I anticipated could potentially exert a notable impact on the overall performance of my predictive model. Specifically, I chose to apply the Multi-Layer Perceptron (MLP) model to both the imbalanced and balanced class datasets in order to comprehensively examine the ramifications of class imbalance on the model's predictive capabilities.

In the ensuing sections, I shall delve into the results obtained from this analysis, which are visually depicted in Figure 1. These results serve as a testament to the significant influence wielded by imbalanced class distributions, as they underscore the marked diminishment in the predictive power and discriminative capacity of the model when confronted with such data disparities.

It is imperative to expand upon this critical observation, as it underscores the paramount importance of addressing class imbalance issues in the context of machine learning. Moreover, I shall elucidate the various methodologies and techniques employed in my study to mitigate the adverse effects of class imbalance, with the overarching goal of enhancing the model's resilience and predictive performance in real-world scenarios.

As indicated by the outcomes presented in the subsequent tables, the following aspects are of paramount significance:

Confusion Matrices:

All four models seem to perform well in correctly classifying the majority of samples across the three classes (-1, 0, 1).

The diagonal elements of the confusion matrices represent the true positives (correctly classified instances) for each class, and off-diagonal elements represent misclassifications.

In terms of misclassifications, it appears that all models have some degree of difficulty distinguishing between classes -1 and 0.

Classification Report Metrics:

Precision: Precision measures the ratio of correctly predicted positive observations to the total predicted positive observations. A higher precision indicates fewer false positives.

Recall: Recall measures the ratio of correctly predicted positive observations to the total actual positive observations. A higher recall indicates fewer false negatives.

F1-score: The F1-score is the harmonic mean of precision and recall and provides a balance between these two metrics.

Now, let's break down the results for each model:

- **SVM:**

Precision and recall values for all classes are decent, but they are slightly lower than those of the Random Forest, XGBoost, and MLP models.

The F1-score for class 0 is lower compared to the other models.

SVM seems to perform reasonably well but might benefit from some hyperparameter tuning to improve performance further.

- **Random Forest:**

The Random Forest model generally performs well, with high precision, recall, and F1-scores for all classes.

It has the highest F1-scores for classes -1 and 1 among the models, indicating good balance between precision and recall.

Random Forest appears to handle class imbalances well and provides solid results.

- **XGBoost:**

XGBoost performs similarly to the Random Forest, with high precision, recall, and F1-scores for all classes.

It has the highest recall for class -1 among the models, indicating it's better at capturing true negatives.

Like Random Forest, XGBoost seems to handle class imbalances effectively.

- **MLP (Multi-Layer Perceptron):**

MLP performs well but has slightly lower precision and recall for class -1 compared to the Random Forest and XGBoost models.

It has a very high F1-score for class 1, indicating excellent performance in capturing true positives for this class.

MLP might need further tuning or architecture adjustments to improve its performance for class 1.

- **Overall Comparisons and Inferences:**

Random Forest and XGBoost consistently provide strong performance across all metrics, making them excellent choices for this classification task.

SVM and MLP also perform well but have some minor weaknesses in terms of precision and recall, especially for class -1.

The choice between Random Forest and XGBoost depends on the specific requirements and the dataset. Both models are strong candidates and are robust to overfitting.

Further hyperparameter tuning, feature engineering, or data preprocessing could potentially improve the performance of all models.

Consideration of the specific business or research goals and the importance of precision and recall for each class should guide the final model selection.

- **Imbalanced Class:**
- **MLP**

  The results of the model are presented in the Table. 7. Also, to show the loss and accuracy of the model per each epoch presented in the Fig. 22.

Table 7. MLP imbalanced Class Results (own research)

Confusion Matrix:

	Predicted -1	Predicted 0	Predicted 1
Actual -1	4	60	0
Actual 0	1	6368	12
Actual 1	0	61	14

Classification Report:

	Precision	Recall	F1-Score	Support
Class -1	0.8	0.06	0.12	64
Class 0	0.98	1	0.99	6381
Class 1	0.54	0.19	0.28	75

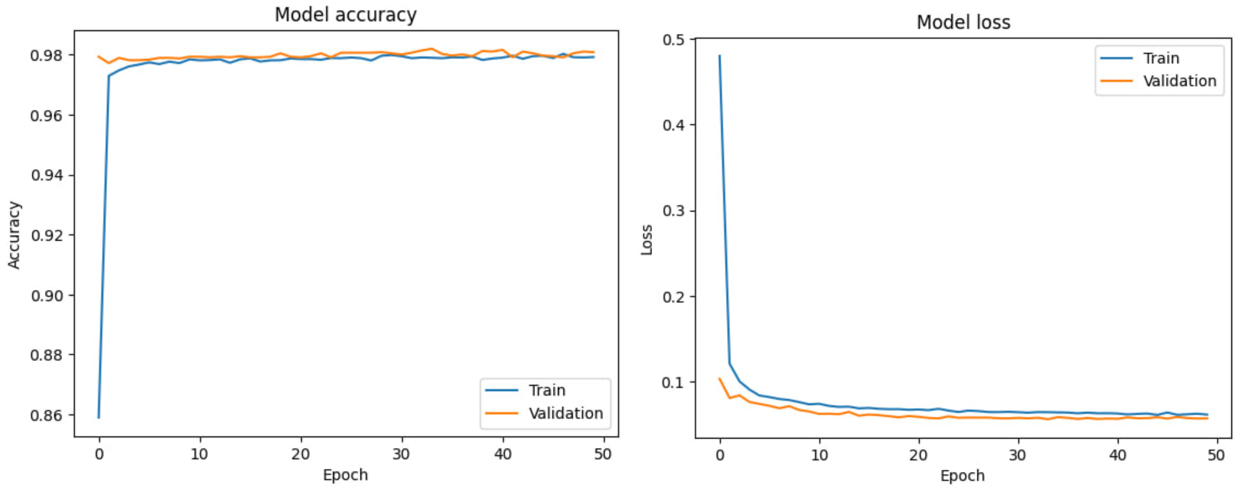


Figure 22. Model accuracy and loss in imbalanced class data (own research)

- **Balanced Class:**
- **MLP**

The results of the model are presented in the Table. 8. Also, to show the loss and accuracy of the model per each epoch presented in the Fig. 23.

Table 8. MLP balanced Class Results (own research)

Confusion Matrix:

	Predicted -1	Predicted 0	Predicted 1
Actual -1	54	14	0
Actual 0	6	92	4
Actual 1	2	0	68

Classification Report:

	Precision	Recall	F1-Score	Support
Class -1	0.87	0.79	0.83	68
Class 0	0.87	0.9	0.88	102
Class 1	0.94	0.97	0.96	70

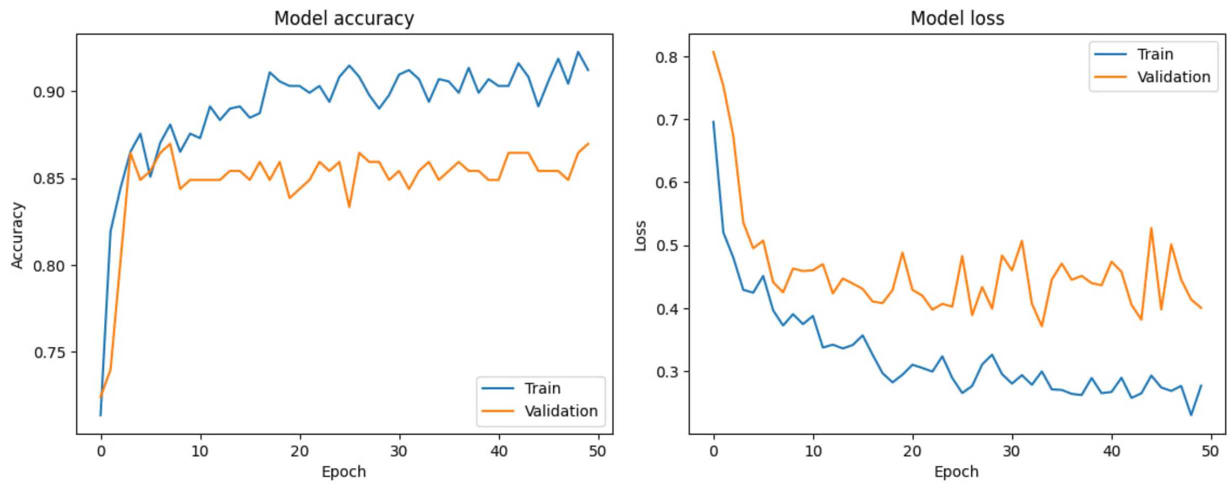


Figure 23. Model accuracy and loss in balanced class data (own research)

- **Balanced Class:**
- **SVM**

The results of the model are presented in the Table 9.

Table 9. SVM balanced Class Results (own research)

Confusion Matrix:

	Predicted -1	Predicted 0	Predicted 1
Actual -1	51	14	3
Actual 0	6	92	4
Actual 1	0	0	70

Classification Report:

	Precision	Recall	F1-Score	Support
Class -1	0.89	0.75	0.82	68
Class 0	0.87	0.9	0.88	102
Class 1	0.91	1	0.95	70

- **Imbalanced Class:**
- **SVM**

The results of the model are presented in the Table. 10.

Table 10. SVM Imbalanced Class Results (own research)

Confusion Matrix:

	Predicted -1	Predicted 0	Predicted 1
Actual -1	0	64	0
Actual 0	0	6381	0
Actual 1	0	75	0

Classification Report:

	Precision	Recall	F1-Score	Support
Class -1	0.00	0.00	0.00	64
Class 0	0.98	1	0.99	6381
Class 1	0.00	0.00	0.00	75

- **Balanced Class:**
- **Random Forest**

The results of the model are presented in the Table. 11.

Table 11. Random Forest balanced Class Results (own research)

Confusion Matrix:

	Predicted -1	Predicted 0	Predicted 1
Actual -1	64	4	0
Actual 0	4	94	4
Actual 1	0	0	70

Classification Report:

	Precision	Recall	F1-Score	Support
Class -1	0.94	0.94	0.94	68
Class 0	0.96	0.92	0.94	102
Class 1	0.95	1	0.97	70

- **Imbalanced Class:**
- **RandomForest**

The results of the model are presented in the Table. 12.

Table 12. Random Forest Imbalanced Class Results (own research)

Confusion Matrix:

	Predicted -1	Predicted 0	Predicted 1
Actual -1	7	57	0
Actual 0	10	6349	22
Actual 1	0	23	52

Classification Report:

	Precision	Recall	F1-Score	Support
Class -1	0.41	0.11	0.17	64
Class 0	0.99	0.99	0.99	6381
Class 1	0.70	0.69	0.70	75

- **Balanced Class:**
- **XGBoost**

The results of the model are presented in the Table. 13.

Table 13. XGBoost balanced Class Results (own research)

Confusion Matrix:

	Predicted -1	Predicted 0	Predicted 1
Actual -1	62	6	0
Actual 0	1	97	4
Actual 1	0	0	70

Classification Report:

	Precision	Recall	F1-Score	Support
Class -1	0.98	0.91	0.95	68
Class 0	0.94	0.95	0.95	102
Class 1	0.95	1	0.97	70

- **Imbalanced Class:**
- **XGBoost**

The results of the model are presented in the Table. 14.

Table 14. XGBoost Imbalanced Class Results (own research)

Confusion Matrix:

	Predicted -1	Predicted 0	Predicted 1
Actual -1	4	60	0
Actual 0	4	6357	20
Actual 1	0	20	55

Classification Report:

	Precision	Recall	F1-Score	Support
Class -1	0.50	0.06	0.11	64
Class 0	0.99	1.00	0.99	6381
Class 1	0.73	0.73	0.73	75

My research has yielded a wealth of insights through a comprehensive examination of various models, providing a nuanced perspective on their respective performance. The data-driven evidence gleaned from my experiments has revealed that, intriguingly, the differences in performance among these models are not starkly pronounced. This revelation prompts us to explore alternative criteria for model selection, transcending the conventional reliance on performance metrics alone.

Within my ensemble of models, one stands out prominently for its outstanding performance – XGBoost, a formidable machine learning model. It not only demonstrates superior predictive capabilities but also distinguishes itself with its remarkable computational efficiency. In comparison, the Multi-Layer Perceptron (MLP), a neural network, lags slightly behind in terms of performance metrics. Notably, XGBoost's computational efficiency renders it particularly appealing for scenarios where rapid results are imperative, underscoring its practical utility.

A pivotal consideration in my model evaluation process is the requisite dataset size. Notably, machine learning models, including XGBoost, exhibit an admirable ability to deliver impressive results even when confronted with relatively modest datasets. In contrast, MLP, given its complex architecture, demands a substantially larger corpus of data to fully unlock its potential. This divergence in data requirements carries significant implications, especially in contexts where access to extensive data resources is limited.

In light of these discerning observations, I have made a judicious selection for my future research endeavors. The decision to embrace XGBoost as my model of choice is underpinned by its formidable combination of robust performance, computational efficiency, and adaptability to datasets of varying magnitudes. This strategic choice aligns with my commitment to extracting optimal insights from my data, ensuring that my research remains both effective and practical.

Furthermore, my adoption of XGBoost extends beyond mere model selection; I leverage its capabilities for the critical tasks of model validation and data testing. Its proficiency in generalizing to previously unseen data bolsters my confidence in its applicability to real-world scenarios, thereby enhancing the reliability and relevance of my research outcomes.

In sum, my methodical and data-driven approach to model evaluation has culminated in the selection of XGBoost as the linchpin of my research efforts. This decision, substantiated by its exemplary performance, computational efficiency, and adaptability to varying dataset sizes, underscores my commitment to achieving the highest possible standards in my research. As I progress with my investigation, the utilization of XGBoost for validation and testing reaffirms its pivotal role in advancing my research objectives.

Tables 15, 16, and 17 present the predicted and actual classes of the model, demonstrating its performance and accuracy.

Table 15. The comparison between the predicted and actual class of selected nodes (Class 1) (own research)

Node	y_actual	y_predicted
check	1	1
performance	1	1
leader	1	1
software development	1	1
system	1	1
build	1	1
advisory	1	1
right	1	1
powerful	1	1
lead	1	1
service	1	1
mobile	1	1
user	1	1
learning	1	1
risk	1	1
budget	1	1
write	1	1
idea	1	1
topic	1	1
methodology	1	1
science	1	1
pro	1	1
global	1	1

Table 16. The comparison between the predicted and actual class of the selected nodes (Class 0) (own research)

Node	y_actual	y_predicted
use	0	0
address	0	0
business want	0	0
british columbia	0	0
basic sap	0	0
arizona	0	-1
automation software	0	0
chart office using	0	0
building new ballistic	0	0
anywhere ecornell	0	0
beginner guide	0	0
best consultation	0	0
aerial tramway funicular	0	0
change business	0	0
ancient	0	0
chapter erp	0	0
anecdote	0	0
better visual	0	0
book experience	0	0
improve	0	0
best linkedin	0	0
blue	0	0
chart	0	0
build successful office	0	0
become ninja	0	0
bachelor	0	0
best tool educational	0	0
better outcome	0	0
airbnb cleaner	0	0
apply business	0	0
area improvement	0	0

Table 17. The comparison between the predicted and actual class of the selected nodes (Class -1) (own research)

Node	y_actual	y_predicted
customer	-1	0
template	-1	-1
training	-1	-1
everything	-1	-1
definition	-1	-1
university	-1	-1
comprehensive	-1	-1
find	-1	-1
future	-1	-1
microsoft	-1	-1
help	-1	-1
successful	-1	-1
skill	-1	-1
ultimate	-1	-1
question	-1	-1
industry	-1	-1
source	-1	-1
real	-1	-1
complete	-1	-1
essential	-1	-1
boost	-1	-1
research	-1	-1

4.11. Validation

In this section, my primary objective was to rigorously assess the efficacy of the models introduced in the preceding section. As previously outlined, I embarked on the practical implementation of an array of diverse machine learning algorithms, specifically encompassing support vector machines (SVM), random forests, and XGBoost. Additionally, I ventured into the realm of deep learning with the inclusion of a multi-layer perceptron (MLP) architecture.

My comprehensive evaluation encompassed a multifaceted approach, involving the meticulous calibration and examination of each model's performance. Through this systematic investigation, I aimed to gauge their respective strengths and weaknesses, thereby providing valuable insights into their suitability for the intended application. This in-depth analysis allowed us to elucidate the robustness and predictive capabilities of these models in a more profound manner, shedding light on their potential utility in practical scenarios.

My approach to model development involved a comprehensive training regimen using a well-structured training dataset. This training phase is critical in honing the models' ability to learn and generalize patterns from the provided data. Following this rigorous training process, I embarked on a meticulous evaluation of the models' performance, a pivotal step in assessing their effectiveness.

To guarantee the dependability of my assessment, I chose a dual data partitioning approach, dividing the dataset into separate portions: the training dataset and the testing dataset. The training dataset acted as the base for constructing my models, with each model acquiring knowledge from the patterns and relationships present in the data.

Subsequently, the test dataset was employed as an independent benchmark to evaluate how well the trained models could generalize to new, unseen data. This partitioning approach mitigated the risk of overfitting, ensuring that the models' performance could be assessed on previously unencountered instances.

During the evaluation phase, I employed a range of diverse evaluation metrics tailored to the specific problem at hand. These metrics allowed us to quantitatively assess various aspects of model performance, including accuracy, precision, recall, F1-score, and more. This multifaceted evaluation approach provided a holistic view of the models' capabilities and their strengths and weaknesses.

Upon comprehensive analysis of the model performance metrics, a notable finding emerged. Specifically, among the suite of machine learning models employed, the random forest and XGBOOST algorithms consistently exhibited the most promising and robust results. These models demonstrated a superior ability to generalize from the training data to the test data, indicative of their capacity to capture underlying patterns in the dataset effectively.

The process of partitioning my dataset into distinct sections and evaluating my model's performance against these partitions represents a critical aspect of my analytical framework. This methodology provides an initial yardstick for assessing the model's competence. However, to substantiate its prowess in terms of generalization and to ensure its consistency in real-world applications, it is imperative to subject the model to a more rigorous evaluation, one that involves entirely unseen data.

In my pursuit of model validation, I adopted a multifaceted approach. The first validation technique involved randomly selecting data points from within the dataset. This approach allowed us to gauge the model's ability to replicate observed patterns within the data it had been trained on. By comparing the actual outcomes with the model's predictions, I gained valuable insights into its performance in scenarios resembling those encountered during training.

My second validation method, which is elaborated upon in the subsequent section, delved into the use of previously unseen data. This approach is of paramount importance as it simulates real-world scenarios where the model encounters data instances that it has not been explicitly exposed to during the training phase. This process of introducing unseen data serves as a litmus test for the model's adaptability and generalization capabilities, offering a robust measure of its potential utility and reliability in practical applications.

To further enhance the depth of my evaluation, I also considered additional factors such as cross-validation techniques, ensuring that my model's performance was rigorously assessed across multiple folds of data partitioning. This additional layer of scrutiny helped us gauge the model's stability and its ability to perform consistently across various subsets of the data.

4.11.1. Random validation

In this section of my study, my objective was to rigorously assess the performance of the model by subjecting it to rigorous testing using the pre-existing dataset. To achieve this goal, I employed a systematic approach involving the random selection of data points from the existing dataset. These selected data points were then utilized as inputs for the model, allowing us to comprehensively evaluate the model's capabilities and robustness.

As depicted in the subsequent tables, the outcomes of my evaluation unequivocally demonstrate the model's exceptional proficiency in the task of classifying nodes. The model exhibited remarkable prowess in discerning patterns and making accurate classifications. These tables serve as a tangible representation of the model's performance, providing a clear and quantitative illustration of its ability to successfully categorize nodes within the dataset.

My rigorous testing methodology, as evidenced by the results presented in these tables, underscores the model's effectiveness and reliability. These findings not only contribute valuable insights into the model's aptitude but also establish a solid foundation for its potential application in practical scenarios, where accurate node classification holds paramount importance.

- **RandomForest**
  The results of the model validation are presented in the Table 18.

Table 18. Random Forest Validation Results (Random Validation) (own research)

Confusion Matrix:

	Predicted -1	Predicted 0	Predicted 1
Actual -1	34	2	0
Actual 0	2	44	1
Actual 1	0	0	37

Classification Report:

	Precision	Recall	F1-Score	Support
Class -1	0.94	0.94	0.94	36
Class 0	0.96	0.94	0.95	47
Class 1	0.97	1.00	0.99	37

- **XGBoost**
  The results of the model validation are presented in the Table. 19.

Table 19. XGBoost Validation Results (Random Validation) (own research)

Confusion Matrix:

	Predicted -1	Predicted 0	Predicted 1
Actual -1	65	5	0
Actual 0	1	135	6
Actual 1	0	0	87

Classification Report:

	Precision	Recall	F1-Score	Support
Class -1	0.98	0.93	0.96	70
Class 0	0.96	0.95	0.96	142
Class 1	0.94	1.00	0.97	87

4.11.2. Unseen Validation

In my pursuit of a comprehensive model assessment, I endeavored to subject the model to the rigors of unseen data evaluation. This approach involved the careful provision of data that the model had not been previously exposed to. Subsequently, I meticulously compared the model's predicted outcomes with the actual data, a process that served as the bedrock of my performance evaluation framework.

As elucidated in the preceding chapter, my data acquisition spanned a significant temporal range, encompassing the period from January 1st, 2021, to June 30th, 2023, thereby encapsulating a total of 30 months' worth of data. To ensure the generation of a robust and comprehensive dataset, I employed a methodical approach by aggregating data into distinct temporal periods. Specifically, I amalgamated data from each three-month interval to constitute a single period, resulting in a total of 10 periods.

This temporal segmentation of data allowed us to observe patterns, trends, and model performance across discrete phases, thereby providing a nuanced understanding of how the model performed under varying conditions. The utilization of 10 distinct periods ensured a diverse and comprehensive dataset, enhancing the model's adaptability and capacity to generalize effectively.

This meticulous data handling approach not only facilitated a more detailed analysis of the model's performance but also allowed us to explore its consistency and robustness across different temporal contexts. It is through this structured methodology that I sought to uncover valuable insights into the model's ability to provide accurate predictions and classifications within a dynamic and evolving dataset.

In the context of my classification task, I employed a rigorous methodology to assess my model's performance over a series of consecutive temporal periods. The central objective was to calculate target values denoted as 1, 0, and -1 for each of these periods. It's noteworthy that the final temporal period in my dataset did not possess these target values; instead, its sole purpose was to label the preceding period, which corresponds to the 1st period of 2023. Consequently, this last period served a unique role as unseen data, challenging the model's predictive capabilities.

To effectively incorporate this period into my evaluation process, I needed to assign target values to each data instance within it. This task demanded the acquisition of a supplementary dataset that would enable us to establish these target values based on degree rankings. Thus, my attention turned to gathering data from subsequent periods, with a specific focus on the 11th period, corresponding to the 3rd period of 2023.

In my quest to acquire this additional dataset, I embarked on a data collection endeavor from online sources. The timeframe covered in this effort extended from July 1st, 2023, to September 30th, 2023, spanning a three-month duration. This dataset played a pivotal role in my evaluation framework, providing the essential information required to assign target labels to the instances within the previously unlabelled period.

The systematic integration of these additional data elements enriched my evaluation process substantially. It enabled us to comprehensively assess the model's predictive capabilities on previously unseen data while highlighting the adaptability and versatility of my analytical framework in addressing complex temporal scenarios. This approach not only ensured a thorough evaluation of the model but also emphasized the importance of dynamic data handling in real-world applications, where temporal dynamics play a crucial role in model performance.

Subsequent to the acquisition of the supplementary dataset encompassing the 11th period, as previously described, my next crucial step involved the utilization of this previously unseen data to rigorously assess the performance of my models. This evaluation process was conducted to gain insights into how effectively the models could extend their predictive capabilities beyond the confines of the training dataset.

The outcomes of this rigorous evaluation are meticulously presented in the subsequent tables, which serve as an invaluable visual representation of my findings. It becomes abundantly evident from these tables that the performance of my models is nothing short of exceptional. The models consistently demonstrated a remarkable aptitude for prediction, effectively capturing and replicating the ranking dynamics of nodes over subsequent time periods.

The impressive performance metrics observed in these tables not only underscore the robustness of my models but also provide compelling evidence of their capacity for generalization. This quality is of paramount importance, as it suggests that the models are capable of making accurate predictions even in scenarios involving data instances they have not encountered during their training phase.

A noteworthy observation that emerges prominently from my results pertains to the distinction between the Random Forest and XGBoost models. This distinction is particularly intriguing as it underscores the nuances of model performance across different phases of my analysis, notably during the training and validation stages.

During the training phase, it is discerned that XGBoost exhibited a marginal advantage in terms of performance metrics. This slight edge suggests that XGBoost was more adept at learning and capturing the underlying patterns present within my training dataset, yielding favorable results in this context.

Conversely, as I transitioned to the validation phase, a notable shift occurred, with Random Forest surpassing XGBoost in terms of performance. This shift highlights the dynamic nature of model performance, as Random Forest demonstrated superior predictive capabilities when faced with unseen data and real-world scenarios, thereby attaining better validation results.

This fluctuation in model performance underscores the significance of comprehensive evaluation. It underscores the notion that a model's proficiency can vary when subjected to distinct phases of the analysis and different subsets of data. This variability in performance between training and validation phases provides valuable insights into the models' adaptability and generalization abilities, serving as a reminder that model assessment should encompass multiple dimensions and real-world conditions.

- **XGBOOST**

  The results of the model validation are presented in the Table. 20

Table 20. XGBoost Validation Results (Unseen Validation) (own research)

Confusion Matrix:

	Predicted -1	Predicted 0	Predicted 1
Actual -1	334	19	0
Actual 0	1	494	5
Actual 1	0	0	342

Classification Report:

	Precision	Recall	F1-Score	Support
Class -1	1	0.95	0.97	353
Class 0	0.96	0.99	0.98	500
Class 1	0.99	1.00	0.99	342

- **Random Forest**
  The results of the model validation are presented in the Table. 21

Table 21. Random Forest Validation Results (Unseen Validation) (own research)

Confusion Matrix:

	Predicted -1	Predicted 0	Predicted 1
Actual -1	347	6	0
Actual 0	4	492	4
Actual 1	0	0	342

Classification Report:

	Precision	Recall	F1-Score	Support
Class -1	0.99	0.98	0.99	353
Class 0	0.99	0.98	0.99	500
Class 1	0.99	1.00	0.99	342

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

The essence of this research originated from the practical and theoretical challenges identified in project management, specifically the overwhelming number of methods and their implications for decision-making. The motivation behind this study stemmed from a desire to streamline these processes, coupled with a passion to contribute to the field. As a project manager turned Ph.D. researcher, I observed the rapid evolution of project management tools and the need for continuous learning to remain relevant. This study acknowledges the necessity for project managers to discriminate between valuable advancements and transient trends. The research journey, although beginning with a broad vision, refined to focus on developing a tool that provides a comprehensive understanding of the current and future landscapes in project management. It incorporates techniques like text mining, text analysis, scientometrics, bibliometrics, and network analysis to navigate and understand the vast and evolving knowledge within the domain, ultimately aiming to forecast future trends and aid both researchers and practitioners.

The current state of project management, while critical, is merely a temporal cross-section of a dynamically changing field. It is foundational, but not all-encompassing, as the tools and techniques that are effective today may become outdated as new technologies, market dynamics, and global events unfold. Similar to finance and meteorology, where understanding present conditions is insufficient for future forecasts, project management requires a predictive gaze that considers historical patterns and external factors. This research, therefore, extends beyond the current paradigm, seeking to equip project managers with the insights necessary to navigate future uncertainties, ensuring they are prepared to adapt and excel in the face of inevitable changes.

In the early phase of my research, I focused on collecting data through a keyword-centric approach centered on 'project management', drawing from both scholarly and practical sources to encapsulate a comprehensive view of the field. This approach facilitated the accumulation of a wide-ranging dataset that informed my network analysis, which aimed to elucidate the complex web of connections and dependencies within the data. my meticulous network mapping allowed us to identify and rank the prominence of various entities and concepts, setting the stage for a more advanced, detailed exploration of the relationships and interactions that define the project management domain.

My exploration of network analysis methodologies was broad and nuanced, incorporating a variety of techniques to unravel different facets of the network. I assessed nodes' importance through Degree Centrality, examined key connectors with Betweenness Centrality, and evaluated nodes' efficiency with Closeness Centrality. Eigenvector Centrality and PageRank Centrality allowed us to identify influential nodes, while the HITS algorithm differentiated between hubs and authoritative nodes. I further examined community structures using the Clustering Coefficient, analyzed connectivity patterns through Assortativity Coefficient, and identified critical communication paths with Load Centrality. Coreness Centrality revealed the network's densest areas, Harmonic Centrality highlighted optimal interconnections, and Average Neighbor Degree provided insight into local neighborhood structures. Finally, Closeness Vitality underscored the nodes integral to the network's efficiency, painting a comprehensive picture of the network's intricate dynamics.

Thomas Kuhn's philosophy revolutionized the perception of scientific progress by introducing the concept of "paradigm shifts" to describe fundamental changes in scientific disciplines. Contrary to the idea of a smooth, continuous advancement, Kuhn proposed that periods of "normal science" operate under a consensus of beliefs and practices until accumulating anomalies trigger a revolutionary phase, leading to a new paradigm. His work highlights the interplay between objective evidence and the subjective consensus of the scientific community, suggesting that scientific truth is, to some extent, socially constructed. Kuhn's ideas have since transcended academic discussions, affecting broader cultural narratives and emphasizing the role of historical and social contexts in scientific development.

The network analysis metrics such as the I-indicator and clustering coefficient (C-indicator) are crucial for assessing the potential and structure within a network. The I-indicator helps identify nodes with significant growth potential by considering the ratio of their betweenness centrality to their degree, thus optimizing network interventions and enhancing predictive models. The C-indicator measures node clustering to understand local structures and identify robust communities within the network. Additionally, the ICCO rank combines these indicators with keyword frequency to gauge a node's overall potential and importance, guiding strategic focus within the network.

Kuhn's stages of scientific development can be related to the use of network analysis indicators like the I-indicator, C-indicator, and ICCO ranking in recognizing and fostering paradigm shifts. In the pre-paradigm phase, these indicators could assess the potential and connectivity of competing theories. During normal science, they could monitor the centrality and clustering of dominant theories. In times of crisis, these tools could identify emerging theories that might lead to a paradigm shift. After a shift, they would help in assessing the new paradigm's growth potential and integration. These indicators, therefore, serve as quantitative tools for analyzing the structure and dynamics of scientific knowledge, aligning with Kuhn's philosophy by potentially signaling the conditions for scientific revolutions.

Machine learning has been a crucial tool in my research, offering a powerful means to analyze and understand patterns within textual data. I began by labeling keywords based on their frequency and prominence in the data, with 'one' indicating a rise in prominence, 'minus one' signifying a drop, and 'zero' for no change in status. However, this approach led to an imbalanced label distribution, which became apparent during the training phase of my machine learning models. I selected four robust algorithms—Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), Random Forest, and GXBoost—for my initial training. The imbalance affected the performance of these models, prompting us to balance my dataset. This recalibration allowed for retraining the models, resulting in a more reliable and insightful analysis.

My methodology is designed to detect shifts in the use of natural language processing terms in project management. I focus on the most frequent 100 keywords, analyzing their prominence over time. A temporal structure is established for comparing keyword significance across quarters, with labels indicating increased, decreased, or stable relevance. This structured approach not only enables us to track changes in keyword importance but also refines the focus of my research. As the investigation continues, I anticipate further refining my methods, enhancing the study's accuracy, and deepening my comprehension of the evolving landscape of project management language.

In my investigation, the comparative performance of various machine learning models revealed only marginal differences. However, XGBoost distinguished itself as the preferred model due to its predictive accuracy and computational efficiency, making it ideal for time-sensitive tasks. It performs well even with smaller datasets, an important consideration when data is scarce. I selected XGBoost for its balanced attributes, confirmed through validation and testing, making it a cornerstone for my research and ensuring its relevance in practical applications.

The validation phase was crucial in assessing the real-world efficacy of my chosen models. I employed a thorough evaluation strategy, using separate training and testing datasets to ensure models could generalize well beyond the data they were trained on. The use of various performance metrics, including accuracy, precision, recall, and F1-score, showed that random forests and XGBoost were the most effective. By testing these models over ten periods and on new, unseen data, I confirmed their robustness and consistent predictive power, solidifying their potential for practical application in varied scenarios.

5.2. This study’s contributions

The study contributes significantly to the field of project management through various innovative approaches and analytical methods:

1. *Identification of Research Gaps:* The study addresses specific practical and theoretical gaps in project management, particularly in decision-making processes, providing the motivation for a thorough investigation.
2. *Comprehensive Data Collection and Network Construction:* It encompasses a meticulous data collection from a wide array of academic and practical sources. A network analysis is constructed using a keyword-centric approach, illustrating the complex interrelations within the project management domain.
3. *Development of Network Analysis Methodologies:* A suite of network analysis techniques is applied, employing diverse centrality measures and tools to dissect the structure and dynamics within the field of project management.
4. *Machine Learning Application:* The research integrates machine learning algorithms for labeling and categorizing keywords, enhancing the understanding of textual data and its significance within project management literature.
5. *Model Selection and Validation Process:* Among various machine learning models analyzed, XGBoost is chosen for its predictive efficiency and computational swiftness, especially beneficial in data-limited contexts. The model's effectiveness is established through a rigorous validation process, ensuring its utility in practical scenarios.
6. *Practical Application and Trend Forecasting:* The study's findings assist project managers in adapting to future changes, utilizing predictive analyses of terminological trends and patterns to prepare for upcoming shifts in the industry.
7. *Incorporation of Kuhn's Scientific Philosophy:* By integrating Thomas Kuhn's philosophy of science with contemporary analytical techniques, the study provides a novel perspective in anticipating paradigm shifts within project management.
8. *Longitudinal Analysis:* A longitudinal approach is adopted, examining data across different time periods, which confirms the consistency and reliability of the chosen models. This aspect is crucial for understanding and forecasting long-term trends within the field.

These contributions underscore the study's role in bridging theoretical frameworks and practical applications, advancing the methodological toolkit available to researchers and practitioners in project management, and paving the way for future scholarly inquiry.

5.3. Future expansion and recommendation:

The foundational work presented in this study offers a robust platform for future expansion in both theoretical and practical realms. Here are potential avenues for advancement:

Theoretical Expansion:

1. *Algorithmic Evolution:* Future research can focus on the development of new algorithms or the refinement of existing ones like XGBoost, tailored specifically for project management data analysis. This could involve machine learning and artificial intelligence to better predict project outcomes.
2. *Model Complexity:* Building more complex models that can take into account the multi-faceted nature of project management, such as stakeholder dynamics, resource allocation, and risk management, could provide deeper insights.

3. *Interdisciplinary Approaches:* Integrating concepts from psychology, sociology, and organizational behavior could enhance understanding of team dynamics and decision-making processes within project management.
4. *Longitudinal Studies:* Further longitudinal studies can be conducted to validate and refine the predictive capabilities of the models over longer periods and across different project types and industries.
5. *Paradigm Shift Analysis:* Expanding on the application of Kuhn's philosophy, researchers could study historical paradigm shifts in project management to better understand and anticipate future changes.

Practical Expansion:

1. *Decision Support Systems:* Development of advanced decision support systems based on the study's findings could assist project managers in real-time decision-making, risk assessment, and strategic planning.
2. *Educational Tools:* The insights from this study can be used to create educational and training programs that focus on the application of machine learning in project management.
3. *Performance Monitoring:* The development of performance monitoring tools that use the study's model to provide ongoing assessment and predictive insights could be invaluable for project managers.
4. *Customization for Industries:* The model can be customized for specific industries or types of projects, providing targeted predictive analytics for sectors such as construction, IT, or healthcare.
5. *Toolkits for Practitioners:* Creation of a toolkit that encapsulates the study's methodologies and findings, offering a user-friendly interface for practitioners to apply these insights in their projects.

Cross-disciplinary Applications:

1. *Beyond Project Management:* The methodologies developed could be tested and adapted for other fields such as supply chain management, logistics, or business process optimization.
2. *Cultural Adaptability:* Future studies could examine how these models perform across different cultural contexts and organizational structures, potentially leading to region-specific models.
3. *Integration with Existing Frameworks:* The models could be integrated with existing project management methodologies like Agile, PRINCE2, or Six Sigma, to enhance their predictive capabilities.

By building upon the groundwork laid by this study, both scholars and practitioners can drive the field of project management toward a more data-driven, predictive, and efficient future. The potential for expansion is significant, with opportunities to make substantial contributions to the theoretical underpinnings of the discipline as well as to the practical tools and techniques used in the industry.

5.4. Research Limitations

Despite the significant contributions of this study, several limitations should be acknowledged. These limitations highlight areas for future research and provide context for interpreting the study's findings.

Data Source Limitations:

- *Data Availability:* The study relied heavily on available data sources, which may not capture the full spectrum of project management practices and trends, especially those emerging in niche or less-documented areas.

- *Data Quality:* Variability in the quality and completeness of the data from different sources could impact the robustness and accuracy of the network analysis and machine learning models.

Methodological Constraints:

- *Algorithm Selection:* While XGBoost was identified as the preferred algorithm, the performance of machine learning models can vary based on data characteristics. Other algorithms might perform better under different conditions or with different datasets.
- *Imbalanced Data:* The initial imbalance in keyword labels affected the performance of the machine learning models, and while this was addressed, residual effects may still influence the outcomes.

Scope of Analysis:

- *Keyword-Centric Approach:* The focus on keyword analysis may overlook broader contextual factors and more nuanced insights that a qualitative approach could provide.
- *Temporal Scope:* The study's longitudinal approach covered specific time periods, which may not fully capture long-term trends or recent rapid changes in project management practices.

Model Generalizability:

- *Context-Specific Findings: The models and findings are tailored to the project* management domain and may not be directly applicable to other fields without significant adaptation.
- *Cultural and Organizational Differences:* The study did not extensively account for how cultural or organizational differences might affect the applicability of the findings across different contexts.

Predictive Limitations:

- *Forecasting Accuracy:* While the study aimed to forecast future trends, the inherent uncertainty and complexity of project management mean that predictions are probabilistic and not definitive.
- *External Influences:* Unanticipated external factors, such as economic shifts, technological breakthroughs, or global events, can dramatically alter the project management landscape, impacting the relevance of the study's predictions.

Technological and Practical Implementation:

- *Tool Development:* The practical application of the developed models into usable tools for project managers is an ongoing process, and their real-world efficacy and user adoption remain to be fully validated.
- *Resource Constraints:* Implementing advanced machine learning models and network analysis tools requires significant computational resources and expertise, which may not be readily available in all organizations.

These limitations underscore the complexity of the research landscape in project management and suggest areas where further investigation and refinement can enhance the study's contributions and practical impact.

5.5. Practical Utilization Recommendations

The insights and methodologies developed in this study offer several practical applications for project managers across various industries. By tailoring the study’s findings to the specific needs and characteristics of different project types, practitioners can maximize the utility of these tools and approaches.

General Recommendations:

- *Integration with Existing Frameworks:* Incorporate the models and network analysis tools with established project management methodologies like Agile, PRINCE2, or Six Sigma to enhance their predictive capabilities and decision-making processes.
- *Customizable Toolkits:* Develop customizable toolkits that encapsulate the study's methodologies, providing user-friendly interfaces for project managers to apply these insights in their projects.

Specific Project Types:

1. *Construction Projects:*
   - *Risk Management:* Utilize predictive models to identify potential risks and optimize resource allocation, ensuring projects stay on schedule and within budget.
   - *Stakeholder Analysis:* Apply network analysis to map and manage stakeholder relationships, ensuring effective communication and collaboration throughout the project lifecycle.
2. *IT Projects:*
   - *Agile Integration:* Enhance Agile frameworks with predictive analytics to anticipate sprint outcomes and adjust backlogs dynamically, improving responsiveness to changes.
   - *Resource Optimization:* Use machine learning algorithms to predict resource needs and optimize the allocation of developers and other critical resources.
3. *Healthcare Projects:*
   - *Regulatory Compliance:* Implement decision support systems that leverage predictive models to ensure compliance with health regulations and standards.
   - *Patient-Centric Approaches:* Employ network analysis to understand and optimize patient pathways and care delivery processes, improving overall project outcomes.
4. *Marketing Projects:*
   - *Campaign Effectiveness:* Use trend forecasting to predict the success of marketing campaigns and adjust strategies in real-time based on emerging data.
   - *Customer Segmentation:* Apply machine learning to segment customer data, tailoring marketing efforts to different audience groups for maximum impact.
5. *Manufacturing Projects:*
   - *Supply Chain Optimization:* Implement predictive models to anticipate supply chain disruptions and optimize inventory management, ensuring smooth production processes.
   - *Quality Control:* Use network analysis to monitor production quality and identify potential issues early, reducing waste and improving product quality.
6. *Research and Development (R&D) Projects:*
   - *Innovation Tracking:* Leverage scientometric and bibliometric analyses to track emerging research trends and prioritize areas with the highest potential for innovation.
   - *Collaboration Networks:* Map collaboration networks to identify key partners and optimize research efforts across different institutions and disciplines.
7. *Public Sector Projects:*
   - *Policy Impact Analysis:* Use predictive models to assess the potential impacts of new policies and programs, aiding in the development of effective public services.
   - *Community Engagement:* Apply network analysis to map community stakeholders and improve engagement strategies, ensuring that projects meet public needs.

8. *Educational Projects:*
   - *Curriculum Development:* Use trend analysis to predict future skill requirements and adjust educational programs accordingly, ensuring that students are well-prepared for the job market.
   - *Student Performance:* Implement machine learning models to identify factors affecting student performance and develop targeted interventions to support at-risk students.

Implementation Strategy:

- *Training and Development:* Provide training programs for project managers on the use of machine learning and network analysis tools, ensuring they can effectively apply these methods in their projects.
- *Software Solutions:* Develop and distribute software solutions that integrate the study's findings, offering project managers easy access to predictive analytics and network analysis capabilities.
- *Continuous Improvement:* Establish feedback loops to continually refine the tools and methodologies based on user feedback and emerging trends, ensuring they remain relevant and effective.

By focusing on these practical applications, project managers can leverage the advanced analytical tools and insights developed in this study to improve project outcomes, drive innovation, and stay ahead of industry trends.

6. NEW SCIENTIFIC RESULTS

This chapter outlines the innovative scientific outcomes derived from this thesis. The results of my research work, along with their implications and relevance to the field, are enumerated below:

1. **Integration of Machine Learning with network analysis:** I confirmed the potency of integrating Machine Learning with network analysis, particularly in the realms of natural language processing and machine learning algorithms. Despite being underutilized, this fusion has yielded significant advancements, particularly in the precise labeling and categorization of keywords. Through the seamless integration of machine learning techniques, the analysis of textual data in project management literature has attained newfound efficiency and precision, culminating in results that are not only more accurate but also more substantively insightful.

2**. Inclusion of Kuhn's Scientific Philosophy:** I demonstrated the profound impact of incorporating Thomas Kuhn's philosophy of science into modern analytical techniques, providing an innovative lens for anticipating paradigm shifts within project management. This integration has enriched the theoretical framework of the study, offering a profound and comprehensive understanding of the intricate dynamics within the field, thus paving the way for more informed decision-making.

3. **Model Choice and Validation:** I established the superiority of XGBoost among various machine learning models, owing to its exceptional predictive efficiency and computational speed. Through meticulous validation, the efficacy of this model has been unequivocally confirmed, showcasing the immense potential of machine learning in amplifying the accuracy and efficiency of data analysis in real-world applications.

4. **Practical Implementation and Trend Prediction:** I validated the practical implications of my research findings by offering invaluable insights for project managers to adeptly navigate future industry shifts. By harnessing the power of predictive analysis on terminological trends and patterns, practitioners can proactively anticipate and strategize for forthcoming changes within the sector. This pragmatic approach underscores the tangible relevance and utility of the study, equipping practitioners with indispensable tools and strategies.

5. **Longitudinal Examination:** I corroborated the efficacy of the longitudinal examination approach employed in this study, characterized by a meticulous examination of data over various time intervals. This methodological rigor is instrumental in ensuring the consistency and reliability of selected models, thereby enabling a comprehensive understanding and forecasting of long-term trends within the field. By incorporating a temporal dimension, the study achieves a nuanced analysis of dynamic shifts, thereby offering valuable insights into future developments in the field of project management.

7. SUMMARY

In this chapter, an attempt has been made to provide a concise overview that encompasses all the steps of the research process leading to the results. The aim is to offer a clear and succinct guide that outlines each stage of the investigation, from the initial hypothesis formation to the final analysis. This comprehensive summary serves as a roadmap, detailing the methodologies employed, the data collected, and the conclusions drawn. It is designed to give readers a complete understanding of the research journey undertaken to arrive at the presented findings. The steps are outlined as follows:

1. Keyword Selection: The research journey began with the careful selection of relevant keywords. This was a critical step as it set the direction for the entire research. The keywords were chosen based on their relevance to the research topic and their ability to yield the most useful results during the web search.

2. Web Search: The selected keywords were then used to conduct a comprehensive search on the web. This wasn't a simple task of entering words into a search engine; it required a strategic approach to ensure that the search results were both relevant and comprehensive. The search was conducted across various databases and online resources to gather as much data as possible related to the research topic. In order to achieve the best results, links are first extracted using the Selenium module. Subsequently, each link is scraped by the Request module. However, it's important to note that the Request module is incapable of conducting searches within a specified time frame.

3. Time Frame Division: To enhance the accuracy of the results, each month was divided into three time frames. This division was based on the understanding that data trends can change significantly within a month. By dividing each month into three-time frames, the research was able to capture these changes and provide a more accurate and detailed analysis of the data.

4. Data Preservation: After the data was collected, it was saved for further analysis. This step was not just about storing data; it was about organizing the data in a way that would make the subsequent steps more efficient. The data was categorized and labeled appropriately to ensure easy retrieval during the analysis phase.

5. Data Preprocessing: Next step involved preprocessing the saved data. This step was crucial in ensuring the quality of the research findings. The preprocessing involved cleaning the data by removing any irrelevant or erroneous data points. It also involved conducting preliminary analyses to identify any patterns or trends in the data. This step served as a bridge between the data collection and the data analysis phases of the research.

6. Network Creation: The research journey continued with the creation of a network. This network served as the initial structure for the research, providing a visual representation of the relationships between different elements. This was not a mere assembly of nodes and edges; it was a carefully constructed framework that encapsulated the complexity of the research topic. The network was created using advanced network creation techniques, ensuring that it accurately represented the relationships between the elements.

7. N-gram Network Creation: The next step was to create different networks based on n-grams. This allowed for a more detailed analysis of the relationships between elements and provided a more nuanced understanding of the data. The n-gram networks were created by considering different combinations of n-grams, which added another layer of depth to the analysis. This step was crucial in uncovering hidden patterns and relationships in the data.

8. Network Analysis: Once the networks were created, they were subjected to rigorous analysis. This analysis provided insights into the structure and characteristics of the networks. The analysis was not limited to simple metrics like node degree or clustering coefficient; it delved deeper into the network structure, uncovering patterns and trends that were not immediately apparent. The network analysis was conducted using advanced network analysis techniques, ensuring that the results were both accurate and insightful.

9. Node Filtering: The top nodes were subsequently sifted based on their degree. This phase aided in pinpointing the most impactful nodes in the network. The filtering was not a mere procedure of choosing the nodes with the utmost degree; it necessitated a meticulous analysis of the nodes' functions and locations within the network. The node filtering process was guided by the research objectives, ensuring that only the most relevant nodes were selected for further analysis.

10. Node Labeling: The nodes were then labeled. This step involved assigning meaningful labels to the nodes, which facilitated the interpretation of the network analysis results. The labeling process was guided by the research questions and objectives, ensuring that the labels were relevant and informative. The process of labeling nodes was predicated on the top 100 nodes and the fluctuations in their rankings over two successive months.

11. Machine Learning Application: Machine learning algorithms were then applied to the analysis and labels. This step involved training various models on the data, which allowed for the prediction and classification of unseen data. The machine learning application was a complex process that involved feature selection, model training, and hyperparameter tuning. The machine learning algorithms were carefully selected to ensure that they were suitable for the data and the research objectives. The final duty of the model was the prediction of the labels.

12. Model Selection: The optimal models were subsequently chosen based on their performance. This stage entailed contrasting the performance of various models and opting for those that yielded the most precise forecasts. The process of model selection was steered by an array of metrics, encompassing accuracy, precision, recall, and F1 score. The model selection process was a critical step in ensuring that the research findings were based on the most accurate and reliable models.

13. Model Validation: The selected models were then validated. This step involved testing the models on a validation set to ensure that they were able to generalize well to unseen data. The validation process was a crucial step in assessing the robustness of the models and their ability to make accurate predictions on new data. The model validation process was conducted using advanced validation techniques, ensuring that the models were thoroughly tested and validated.

14. Model Testing: The final step involved testing the models. This step provided a final check on the performance of the models and ensured that they were ready for deployment. The testing process involved applying the models to a test set and evaluating their performance. The model testing process was a rigorous process that ensured that the models were not only accurate but also robust and reliable.

APPENDIX: BIBLIOGRAPHY

1. A Kassem, M, Khoiry, MA & Hamzah, N 2020, ‘Theoretical review on critical risk factors in oil and gas construction projects in Yemen’, *Engineering, Construction and Architectural Management*, vol. 28, Emerald Group Holdings Ltd., no. 4, pp. 934–968.
2. Abbasi, A, Altmann, J & Hossain, L 2011, ‘Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures’, *Journal of Informetrics*, vol. 5, no. 4, pp. 594–607.
3. Abdelkhalik, HF & Azmy, HH 2022, ‘The role of project management in the success of green building projects: Egypt as a case study’, *Journal of Engineering and Applied Science*, vol. 69, Springer Science and Business Media B.V., no. 1, pp. 1–17, viewed 20 April 2023, <https://jeas.springeropen.com/articles/10.1186/s44147-022-00112-5>.
4. AbdiKhalife, MA, Dunay, A & Illés, CB 2021, ‘Bibliometric Analysis of Articles on Project Management Research’, *Periodica Polytechnica Social and Management Sciences*, vol. 29, Budapest University of Technology and Economics, no. 1, pp. 70–83, viewed 7 May 2023, <https://pp.bme.hu/so/article/view/15717>.
5. Abioye, SO, Oyedele, LO, Akanbi, L, Ajayi, A, Davila Delgado, JM, Bilal, M, Akinade, OO & Ahmed, A 2021, ‘Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges’, *Journal of Building Engineering*, vol. 44, Elsevier, p. 103299.
6. Adhikari, K & Poudyal, L 2021, ‘Future of Construction Industry: COVID-19 and Its Implications on Construction Projects and Risk Management – A Review’, Preprints, viewed 4 May 2023, <https://www.preprints.org/manuscript/202104.0383/v1>.
7. Adler, DG, Hilden, K, Wills, JC, Quinney, E & Fang, JC 2010, ‘What drives US gastroenterology fellows to pursue academic vs. non-academic careers?: Results of a national survey’, *American Journal of Gastroenterology*, vol. 105, no. 6, pp. 1220–1223, viewed 16 May 2023, <https://journals.lww.com/ajg/Fulltext/2010/06000/What_Drives_US_Gastroenterology_Fellows_to_Pursue.1.aspx>.
8. ‘Agile or Traditional Project Organization: A Quantitative Assessment of Decision Criteria among Firms in the DACH Region’ 2022, *Central European Business Review*, vol. 11, Vysoká škola ekonomická v Praze - Fakulta podnikohospodářská;, no. 5, pp. 67–83.
9. *Agile Transformation: Structures, Processes and Mindsets for the Digital Age - Neil Perkin - Google Books* n.d., viewed 20 April 2023, <https://books.google.fr/books?hl=en&lr=&id=A5avEAAAQBAJ&oi=fnd&pg=PP1&dq=Agile+project+management+methodologies+like+Scrum+and+Kanban+have+emerged+to+help+manage+fast-moving+projects+with+changing+requirements,+while+virtual+project+management+tools+have+made+it+possible+to+coordinate+teams+across+time+zones+and+geographies.+&ots=uznGbTgzsc&sig=jQGpia8xRGAFwbb__z-kAgopopg&redir_esc=y#v=onepage&q&f=false>.
10. Ahmed, AT, Gohary, F El, Tzanakakis, VA & Angelakis, AN 2020, ‘Egyptian and Greek Water Cultures and Hydro-Technologies in Ancient Times’, *Sustainability 2020, Vol. 12, Page 9760*, vol. 12, Multidisciplinary Digital Publishing Institute, no. 22, p. 9760, viewed 4 May 2023, <https://www.mdpi.com/2071-1050/12/22/9760/htm>.
11. Ahmed, H, Traore, I, Saad, S & Mamun, M 2021, ‘Automated detection of unstructured context-dependent sensitive information using deep learning’, *Internet of Things*, vol. 16, Elsevier, p. 100444.
12. Ahmed, M, Choudhury, S & Al-Turjman, F 2019, ‘Big Data Analytics for Intelligent Internet of Things’, Springer, Cham, pp. 107–127, viewed 7 May 2023, <https://link.springer.com/chapter/10.1007/978-3-030-04110-6_6>.
13. Ahmed, R, Philbin, SP & Cheema, F e. A 2021, ‘Systematic literature review of project manager’s leadership competencies’, *Engineering, Construction and Architectural Management*, vol. 28, Emerald Group Holdings Ltd., no. 1, pp. 1–30.
14. Ajah, IA & Nweke, HF 2019, ‘Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications’, *Big Data and Cognitive Computing 2019, Vol. 3, Page 32*, vol. 3, Multidisciplinary

Digital Publishing Institute, no. 2, p. 32, viewed 7 May 2023, <https://www.mdpi.com/2504-2289/3/2/32/htm>.
15. Akpan, IJ & Ezeume, IC 2021, 'Four Decades Bibliometric Analysis, Science Mapping, and Visualization of the Consequences of Marital Union Dissolution on Parents and Children', *https://doi.org/10.1080/10502556.2021.1993009*, vol. 63, Routledge, no. 1, pp. 1–34, viewed 7 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/10502556.2021.1993009>.
16. Albuquerque, F, Torres, AS & Berssaneti, FT 2020, 'Lean product development and agile project management in the construction industry', *Revista de Gestao*, vol. 27, Emerald Group Holdings Ltd., no. 2, pp. 135–151.
17. Alfreahat, D & Sebestyén, Z 2022, 'A construction-specific extension to a standard project risk management process', *Organization, Technology and Management in Construction*, vol. 14, De Gruyter Open Ltd, no. 1, pp. 2666–2674.
18. Ali, KN, Alhajlah, HH & Kassem, MA 2022, 'Collaboration and Risk in Building Information Modelling (BIM): A Systematic Literature Review', *Buildings 2022, Vol. 12, Page 571*, vol. 12, Multidisciplinary Digital Publishing Institute, no. 5, p. 571, viewed 12 May 2023, <https://www.mdpi.com/2075-5309/12/5/571/htm>.
19. AlMaazmi, JJ 2022, 'A Review And Comparison of Strategy Diffusion (Top-Down) and Performance (Bottom-Up) in Project-Based Organisations', *Creative Construction e-Conference 2022*, Budapest University of Technology and Economics, pp. 112–137, viewed 4 May 2023, <https://repozitorium.omikk.bme.hu/handle/10890/31752>.
20. Al-Makhadmeh, Z & Tolba, A 2020, 'Automatic hate speech detection using killer natural language processing optimizing ensemble deep learning approach', *Computing*, vol. 102, Springer, no. 2, pp. 501–522, viewed 10 May 2023, <https://link.springer.com/article/10.1007/s00607-019-00745-0>.
21. Almanaseer, AM, Alzyadat, W, Muhairat, M, Al-Showarah, S & Alhroob, A 2022, 'A proposed model for eliminating nonfunctional requirements in Agile Methods using natural language processes', *2022 International Conference on Emerging Trends in Computing and Engineering Applications, ETCEA 2022 - Proceedings*, Institute of Electrical and Electronics Engineers Inc.
22. Almazmomi, N, Ilmudeen, A & Qaffas, AA 2022, 'The impact of business analytics capability on data-driven culture and exploration: achieving a competitive advantage', *Benchmarking*, vol. 29, Emerald Group Holdings Ltd., no. 4, pp. 1264–1283.
23. Alryalat, SAS, Malkawi, LW & Momani, SM 2019, 'Comparing Bibliometric Analysis Using PubMed, Scopus, and Web of Science Databases', *JoVE (Journal of Visualized Experiments)*, vol. 2019, Journal of Visualized Experiments, no. 152, p. e58494, viewed 13 May 2023, <https://www.jove.com/v/58494/comparing-bibliometric-analysis-using-pubmed-scopus-web-science>.
24. Alsharari, AS, Wan Zainon, WMN, Letchmunan, S, Mohammed, BA & Alsharari, MS 2023, 'A Review of Agile Methods for Requirement Change Management in Web Engineering', *International Conference on Smart Computing and Application, ICSCA 2023*, Institute of Electrical and Electronics Engineers Inc.
25. AL-Smadi, H, Al-Sakkaf, A, Zayed, T & Nasiri, F 2023, 'An integrated space-based building maintenance management model using multi-objective optimization', *Smart and Sustainable Built Environment*, vol. 12, Emerald Publishing, no. 2, pp. 277–297.
26. Alvesson, M & Deetz, S 2011, 'Doing Critical Management Research', *Doing Critical Management Research*, SAGE Publications Ltd.
27. Alvesson, M & Sandberg, J n.d., 'Pre-understanding: An interpretation-enhancer and horizon-expander in research', *Organization Studies*, vol. 2022, no. 3, pp. 395–412, viewed 16 May 2023, <https://doi.org/10.1177/0170840621994507>.
28. Alvesson, M & Willmott, H 2012, 'Making sense of management: A critical introducton', *Making Sense of Management: A Critical Introducton*, SAGE Publications Inc., pp. 1–308.
29. Alzoubi, HM & Aziz, R 2021, 'Does Emotional Intelligence Contribute to Quality of Strategic Decisions? The Mediating Role of Open Innovation', *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 7, Elsevier, no. 2, p. 130.

30. Ambilkar, P, Dohale, V, Gunasekaran, A & Bilolikar, V 2021, 'Product returns management: a comprehensive review and future research agenda', *https://doi.org/10.1080/00207543.2021.1933645*, vol. 60, Taylor & Francis, no. 12, pp. 3920–3944, viewed 12 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/00207543.2021.1933645>.
31. Amir, A, khodeir, L & Khaled, A 2023, 'Identification of key Risks in Fast-Track Construction projects: A Literature Review', October university for modern sciences and Arts MSA, viewed 4 May 2023, <http://repository.msa.edu.eg/xmlui/handle/123456789/5527>.
32. Amirtash, P, Parchami Jalal, M & Jelodar, MB 2021, 'Integration of project management services for International Engineering, Procurement and Construction projects', *Built Environment Project and Asset Management*, vol. 11, Emerald Group Holdings Ltd., no. 2, pp. 330–349.
33. Amur, ZH, Kwang Hooi, Y, Bhanbhro, H, Dahri, K & Soomro, GM 2023, 'Short-Text Semantic Similarity (STSS): Techniques, Challenges and Future Perspectives', *Applied Sciences 2023, Vol. 13, Page 3911*, vol. 13, Multidisciplinary Digital Publishing Institute, no. 6, p. 3911, viewed 10 May 2023, <https://www.mdpi.com/2076-3417/13/6/3911/htm>.
34. An, L, Grimm, V, Sullivan, A, TurnerII, BL, Malleson, N, Heppenstall, A, Vincenot, C, Robinson, D, Ye, X, Liu, J, Lindkvist, E & Tang, W 2021, 'Challenges, tasks, and opportunities in modeling agent-based complex systems', *Ecological Modelling*, vol. 457, Elsevier, p. 109685.
35. Andrade, PA de, Vanhoucke, M & Martens, A 2023, 'An empirical project forecasting accuracy framework using project regularity', *Annals of Operations Research 2023*, Springer, pp. 1–21, viewed 6 May 2023, <https://link.springer.com/article/10.1007/s10479-023-05269-7>.
36. Aquino, JF, Riss, RR, Multerer, SM, Mogilner, LN & Turner, TL 2022, 'A step-by-step guide for mentors to facilitate team building and communication in virtual teams', *https://doi.org/10.1080/10872981.2022.2094529*, vol. 27, Taylor & Francis, no. 1, viewed 4 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/10872981.2022.2094529>.
37. Arnold, M & Bashir, T 2020, 'Agile Methodologies as a Management Tool for Physical Systems Engineering Development'.
38. Arosio, M, Martina, MLV & Figueiredo, R 2020, 'The whole is greater than the sum of its parts: A holistic graph-based assessment approach for natural hazard risk of complex systems', *Natural Hazards and Earth System Sciences*, vol. 20, Copernicus GmbH, no. 2, pp. 521–547.
39. Asbjornsen, H, Wang, Y, Ellison, D, Ashcraft, CM, Atallah, SS, Jones, K, Mayer, A, Altamirano, M & Yu, P 2022, 'Multi-Targeted payments for the balanced management of hydrological and other forest ecosystem services', *Forest Ecology and Management*, vol. 522, Elsevier, p. 120482.
40. Ashaari, MA, Singh, KSD, Abbasi, GA, Amran, A & Liebana-Cabanillas, FJ 2021, 'Big data analytics capability for improved performance of higher education institutions in the Era of IR 4.0: A multi-analytical SEM & ANN perspective.', *Technological Forecasting and Social Change*, vol. 173, North-Holland, p. 121119.
41. Ashkanani, S & Franzoi, R 2022, 'An overview on megaproject management systems', *Management Matters*, vol. 19, Emerald Publishing Limited, no. 2, pp. 129–148.
42. Asoh, D, Belardo, S & Neilson, R 2002, 'Knowledge management: Issues, challenges and opportunities for governments in the new economy', *Proceedings of the Annual Hawaii International Conference on System Sciences*, vol. 2002-January, IEEE Computer Society, pp. 1745–1754.
43. Assale, M, Dui, LG, Cina, A, Seveso, A & Cabitza, F 2019, 'The Revival of the Notes Field: Leveraging the Unstructured Content in Electronic Health Records', *Frontiers in Medicine*, vol. 6, Frontiers Media S.A., p. 66.
44. Ayat, M, Imran, M, Ullah, A & Kang, CW 2021, 'Current trends analysis and prioritization of success factors: a systematic literature review of ICT projects', *International Journal of Managing Projects in Business*, vol. 14, Emerald Group Holdings Ltd., no. 3, pp. 652–679.
45. Azad, P, Navimipour, NJ, Rahmani, AM & Sharifi, A 2020, 'The role of structured and unstructured data managing mechanisms in the Internet of things', *Cluster Computing*, vol. 23, Springer, no. 2, pp. 1185–1198, viewed 7 May 2023, <https://link.springer.com/article/10.1007/s10586-019-02986-2>.

46. Bag, S, Wood, LC, Xu, L, Dhamija, P & Kayikci, Y 2020, 'Big data analytics as an operational excellence approach to enhance sustainable supply chain performance', *Resources, Conservation and Recycling*, vol. 153, Elsevier, p. 104559.
47. Bailey, AL, Blackstock-Bernstein, A, Ryan, E & Pitsoulakis, D 2016, 'Data mining with natural language processing and corpus linguistics: Unlocking access to school children's language in diverse contexts to improve instructional and assessment practices', *Data Mining And Learning Analytics: Applications in Educational Research*, Wiley Blackwell, pp. 255–275.
48. Baltaru, RD 2018, 'Do non-academic professionals enhance universities' performance? Reputation vs. organisation', *https://doi.org/10.1080/03075079.2017.1421156*, vol. 44, Routledge, no. 7, pp. 1183–1196, viewed 16 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/03075079.2017.1421156>.
49. Bamber, N, Turner, I, Arulnathan, V, Li, Y, Zargar Ershadi, S, Smart, A & Pelletier, N 2020, 'Comparing sources and analysis of uncertainty in consequential and attributional life cycle assessment: review of current practice and recommendations', *International Journal of Life Cycle Assessment*, vol. 25, Springer, no. 1, pp. 168–180, viewed 13 May 2023, <https://link.springer.com/article/10.1007/s11367-019-01663-1>.
50. Baraliuc, I, Depreeuw, S & Gutwirth, S 2013, 'Copyright enforcement in the digital age: a post-ACTA view on the balancing of fundamental rights', *International Journal of Law and Information Technology*, vol. 21, Oxford Academic, no. 1, pp. 92–104, viewed 16 May 2023, <https://academic.oup.com/ijlit/article/21/1/92/682998>.
51. Barnett, MP 1963, 'Continued operation notation for symbol manipulation and array processing', *Communications of the ACM*, vol. 6, no. 8, pp. 467–472.
52. Battisti, S, Agarwal, N & Brem, A 2022, 'Creating new tech entrepreneurs with digital platforms: Meta-organizations for shared value in data-driven retail ecosystems', *Technological Forecasting and Social Change*, vol. 175, North-Holland, p. 121392.
53. Beermann, K & Austin, MC 2021, 'An Inspection of the Life Cycle of Sustainable Construction Projects: Towards a Biomimicry-Based Road Map Integrating Circular Economy', *Biomimetics 2021, Vol. 6, Page 67*, vol. 6, Multidisciplinary Digital Publishing Institute, no. 4, p. 67, viewed 4 May 2023, <https://www.mdpi.com/2313-7673/6/4/67/htm>.
54. Benzidia, S, Makaoui, N & Bentahar, O 2021, 'The impact of big data analytics and artificial intelligence on green supply chain process integration and hospital environmental performance', *Technological Forecasting and Social Change*, vol. 165, North-Holland, p. 120557.
55. De Bernardi, P, Bertello, A, Forliano, C & Orlandi, LB 2021, 'Beyond the "ivory tower". Comparing academic and non-academic knowledge on social entrepreneurship', *International Entrepreneurship and Management Journal 2021 18:3*, vol. 18, Springer, no. 3, pp. 999–1032, viewed 16 May 2023, <https://link.springer.com/article/10.1007/s11365-021-00783-1>.
56. Berrone, P, Rousseau, HE, Ricart, JE, Brito, E & Giuliodori, A 2023, 'How can research contribute to the implementation of sustainable development goals? An interpretive review of SDG literature in management', *International Journal of Management Reviews*, vol. 25, John Wiley & Sons, Ltd, no. 2, pp. 318–339, viewed 4 May 2023, <https://onlinelibrary.wiley.com/doi/full/10.1111/ijmr.12331>.
57. Best, BJ & Lewis, JRT 2007, 'Program Evaluation and Review Technique (PERT) and Critical Path Method (CPM)', *Encyclopedia of Public Administration and Public Policy*, Routledge, pp. 1583–1587, viewed 20 April 2023, <https://www.taylorfrancis.com/chapters/edit/10.1081/e-epap2-324/program-evaluation-review-technique-pert-critical-path-method-cpm-bradley-best-jeremy-lewis>.
58. Bhat, SA & Huang, NF 2021, 'Big Data and AI Revolution in Precision Agriculture: Survey and Challenges', *IEEE Access*, vol. 9, Institute of Electrical and Electronics Engineers Inc., pp. 110209–110222.
59. Bhattacharya, S 2021, 'A Primer on Machine Learning in Subsurface Geosciences', SpringerBriefs in Petroleum Geoscience & Engineering, Springer International Publishing, Cham, viewed 7 May 2023, <https://link.springer.com/10.1007/978-3-030-71768-1>.

60. Bian, R & Tolford, T 2022, ‘Evaluating the Implementation of the Complete Streets Policy in Louisiana: A Review of Practices and Projects in the Last 10 Years’, *https://doi.org/10.1177/03611981221115726*, SAGE PublicationsSage CA: Los Angeles, CA, p. 036119812211157, viewed 4 May 2023, <https://journals.sagepub.com/doi/abs/10.1177/03611981221115726>.
61. Biggi, G & Stilgoe, J 2021, ‘Artificial Intelligence in Self-Driving Cars Research and Innovation: A Scientometric and Bibliometric Analysis’, *SSRN Electronic Journal*, Elsevier BV, viewed 7 May 2023, <https://papers.ssrn.com/abstract=3829897>.
62. Birkle, C, Pendlebury, DA, Schnell, J & Adams, J 2020, ‘Web of Science as a data source for research on scientific and scholarly activity’, *Quantitative Science Studies*, vol. 1, MIT Press, no. 1, pp. 363–376, viewed 13 May 2023, <https://direct.mit.edu/qss/article/1/1/363/15569/Web-of-Science-as-a-data-source-for-research-on>.
63. Bravo, G, Grimaldo, F, López-Iñesta, E, Mehmani, B & Squazzoni, F 2019, ‘The effect of publishing peer review reports on referee behavior in five scholarly journals’, *Nature Communications 2019 10:1*, vol. 10, Nature Publishing Group, no. 1, pp. 1–8, viewed 16 May 2023, <https://www.nature.com/articles/s41467-018-08250-2>.
64. Breyter, M 2022, ‘The Role of Project and Product Management in Software Delivery and IT Services’, *Agile Product and Project Management*, Apress, Berkeley, CA, pp. 3–22, viewed 20 April 2023, <https://link.springer.com/chapter/10.1007/978-1-4842-8200-7_1>.
65. Britto, MT, Fuller, SC, Kaplan, HC, Kotagal, U, Lannon, C, Margolis, PA, Muething, SE, Schoettker, PJ & Seid, M 2018, ‘Using a network organisational architecture to support the development of Learning Healthcare Systems’, *BMJ Quality & Safety*, vol. 27, BMJ Publishing Group Ltd, no. 11, pp. 937–946, viewed 4 May 2023, <https://qualitysafety.bmj.com/content/27/11/937>.
66. Bühler, MM, Nübel, K, Jelinek, T, Riechert, D, Bauer, T, Schmid, T & Schneider, M 2023, ‘Data Cooperatives as a Catalyst for Collaboration, Data Sharing and the Digital Transformation of the Construction Sector’, *Buildings*, vol. 13, MDPI, no. 2.
67. Butera, F 2020, ‘Sociotechnical systems design revisited at the end of the 20th century : STS 2.0.’, *Sociotechnical systems design revisited at the end of the 20th century : STS 2.0.*, Franco Angeli, no. 1, pp. 102–125.
68. Čabarkapa, J 2019, ‘Analysis and comparison of ISO 21500 - Guidance on project management and PMBOK 6th Guide’, Atlantis Press, pp. 266–271, viewed 20 April 2023, <https://www.atlantis-press.com/proceedings/senet-19/125925995>.
69. Camacho, D, Panizo-LLedot, Á, Bello-Orgaz, G, Gonzalez-Pardo, A & Cambria, E 2020, ‘The four dimensions of social network analysis: An overview of research methods, applications, and software tools’, *Information Fusion*, vol. 63, Elsevier, pp. 88–120.
70. Cao, S, Liu, X & Er, H 2010, ‘Dujiangyan Irrigation System – a world cultural heritage corresponding to concepts of modern hydraulic science’, *Journal of Hydro-environment Research*, vol. 4, Elsevier, no. 1, pp. 3–13.
71. Cappa, F, Oriani, R, Peruffo, E & McCarthy, I 2021, ‘Big Data for Creating and Capturing Value in the Digitalized Environment: Unpacking the Effects of Volume, Variety, and Veracity on Firm Performance*’, *Journal of Product Innovation Management*, vol. 38, John Wiley & Sons, Ltd, no. 1, pp. 49–67, viewed 7 May 2023, <https://onlinelibrary.wiley.com/doi/full/10.1111/jpim.12545>.
72. Chairani, C & Siregar, SV 2021, ‘The effect of enterprise risk management on financial performance and firm value: the role of environmental, social and governance performance’, *Meditari Accountancy Research*, vol. 29, Emerald Group Holdings Ltd., no. 3, pp. 647–670.
73. Chandrachooodan, G, Radhika, R & Palappan, RR 2023, ‘Identifying the relevant project management tools in implementation of e-governance projects – Journey from traditional to agile’, *AIP Conference Proceedings*, vol. 2690, AIP Publishing, no. 1, p. 020038, viewed 20 April 2023, <https://pubs.aip.org/aip/acp/article/2690/1/020038/2880322/Identifying-the-relevant-project-management-tools>.
74. Chang, HH & Mukherjee, A 1AD, ‘Using Machine Learning to Extract Insights From Consumer Data’, *https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-7998-9220-*

5.ch107, IGI Global, pp. 1779–1793, viewed 7 May 2023, <https://www.igi-global.com/chapter/using-machine-learning-to-extract-insights-from-consumer-data/317584>.
75. Chen, C, Judge, J & Hulse, D 2022, 'PyLUSAT: An open-source Python toolkit for GIS-based land use suitability analysis', *Environmental Modelling and Software*, vol. 151, Elsevier Ltd.
76. Chen, CK, Palma, F & Reyes, L 2019, 'Reducing global supply chains' waste of overproduction by using lean principles: A conceptual approach', *International Journal of Quality and Service Sciences*, vol. 11, Emerald Group Holdings Ltd., no. 4, pp. 441–454.
77. Cheng, X, Liu, S, Sun, X, Wang, Z, Zhou, H, Shao, Y & Shen, H 2021, 'Combating emerging financial risks in the big data era: A perspective review', *Fundamental Research*, vol. 1, Elsevier, no. 5, pp. 595–606.
78. Chng, SY, Paul, |, Tern, JW, Matthew, |, Kan, RX, Lionel, | & Cheng, TE 2023, 'Automated labelling of radiology reports using natural language processing: Comparison of traditional and newer methods', *Health Care Science*, vol. 2, John Wiley & Sons, Ltd, no. 2, pp. 120–128, viewed 10 May 2023, <https://onlinelibrary.wiley.com/doi/full/10.1002/hcs2.40>.
79. Chofreh, AG, Goni, FA, Klemeš, JJ, Seyed Moosavi, SM, Davoudi, M & Zeinalnezhad, M 2021, 'Covid-19 shock: Development of strategic management framework for global energy', *Renewable and Sustainable Energy Reviews*, vol. 139, Pergamon, p. 110643.
80. Ciric, D, Lalic, B, Gracanin, D, Tasic, N, Delic, M & Medic, N 2019, 'Agile vs. Traditional Approach in Project Management: Strategies, Challenges and Reasons to Introduce Agile', *Procedia Manufacturing*, vol. 39, Elsevier, pp. 1407–1414.
81. Colombo, MG & Piva, E 2012, 'Firms' genetic characteristics and competence-enlarging strategies: A comparison between academic and non-academic high-tech start-ups', *Research Policy*, vol. 41, North-Holland, no. 1, pp. 79–92.
82. *Copyright enforcement in the digital age: a post-ACTA view on the balancing of fundamental rights * | International Journal of Law and Information Technology | Oxford Academic* n.d., viewed 16 May 2023, <https://academic.oup.com/ijlit/article-abstract/21/1/92/682998>.
83. Correia, RA, Ladle, R, Jarić, I, Malhado, ACM, Mittermeier, JC, Roll, U, Soriano-Redondo, A, Veríssimo, D, Fink, C, Hausmann, A, Guedes-Santos, J, Vardi, R & Di Minin, E 2021, 'Digital data sources and methods for conservation culturomics', *Conservation Biology*, vol. 35, John Wiley & Sons, Ltd, no. 2, pp. 398–411, viewed 7 May 2023, <https://onlinelibrary.wiley.com/doi/full/10.1111/cobi.13706>.
84. Cusumano, MA, Holweg, M, Howell, J, Netland, T, Shah, R, Shook, J, Ward, P & Womack, J 2021, 'Commentaries on "The Lenses of Lean"', *Journal of Operations Management*, vol. 67, John Wiley & Sons, Ltd, no. 5, pp. 627–639, viewed 6 May 2023, <https://onlinelibrary.wiley.com/doi/full/10.1002/joom.1138>.
85. Dahl, C 2009, 'Undergraduate research in the public domain: The evaluation of non-academic sources online', *Reference Services Review*, vol. 37, Emerald Group Publishing Limited, no. 2, pp. 155–163.
86. Dai, Z, Yang, Z, Yang, Y, Carbonell, J, Le, Q V. & Salakhutdinov, R 2020, 'Transformer-XL: Attentive language models beyond a fixed-length context', *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, Association for Computational Linguistics (ACL), pp. 2978–2988.
87. Darko, A, Chan, APC, Yang, Y & Tetteh, MO 2020, 'Building information modeling (BIM)-based modular integrated construction risk management – Critical survey and future needs', *Computers in Industry*, vol. 123, Elsevier, p. 103327.
88. Dash, S, Shakyawar, SK, Sharma, M & Kaushik, S 2019, 'Big data in healthcare: management, analysis and future prospects', *Journal of Big Data*, vol. 6, SpringerOpen, no. 1, pp. 1–25, viewed 7 May 2023, <https://link.springer.com/articles/10.1186/s40537-019-0217-0>.
89. Davidavičiene, V, Al Majzoub, K & Meidute-Kavaliauskiene, I 2020, 'Factors Affecting Knowledge Sharing in Virtual Teams', *Sustainability 2020, Vol. 12, Page 6917*, vol. 12, Multidisciplinary Digital Publishing Institute, no. 17, p. 6917, viewed 5 May 2023, <https://www.mdpi.com/2071-1050/12/17/6917/htm>.

90. Debrah, C, Chan, APC & Darko, A 2022, ‘Green finance gap in green buildings: A scoping review and future research needs’, *Building and Environment*, vol. 207, Pergamon, p. 108443.
91. Ding, Y, Ma, J & Luo, X 2022, ‘Applications of natural language processing in construction’, *Automation in Construction*, vol. 136, Elsevier, p. 104169.
92. Dogra, V, Verma, S, Kavita, Chatterjee, P, Shafi, J, Choi, J & Ijaz, MF 2022, ‘A Complete Process of Text Classification System Using State-of-the-Art NLP Models’, *Computational Intelligence and Neuroscience*, vol. 2022, Hindawi Limited.
93. Duan, P, Wang, Y & Yin, P 2020, ‘Remote Sensing Applications in Monitoring of Protected Areas: A Bibliometric Analysis’, *Remote Sensing 2020, Vol. 12, Page 772*, vol. 12, Multidisciplinary Digital Publishing Institute, no. 5, p. 772, viewed 7 May 2023, <https://www.mdpi.com/2072-4292/12/5/772/htm>.
94. Eckert, T & Hüsig, S 2021, ‘Innovation portfolio management: a systematic review and research agenda in regards to digital service innovations’, *Management Review Quarterly 2021 72:1*, vol. 72, Springer, no. 1, pp. 187–230, viewed 4 May 2023, <https://link.springer.com/article/10.1007/s11301-020-00208-3>.
95. Egbelakin, T, Ogunmakinde, OE, Teshich, B & Omotayo, T 2021, ‘Managing Fast-Track Construction Project in Qatar: Challenges and Opportunities’, *Buildings 2021, Vol. 11, Page 640*, vol. 11, Multidisciplinary Digital Publishing Institute, no. 12, p. 640, viewed 4 May 2023, <https://www.mdpi.com/2075-5309/11/12/640/htm>.
96. ElAtia, S, Ipperciel, D & Zaïane, OR 2016, ‘Data Mining And Learning Analytics: Applications in Educational Research’, *Data Mining And Learning Analytics: Applications in Educational Research*, Wiley Blackwell, pp. 1–283.
97. Elghaish, F, Hosseini, MR, Matarneh, S, Talebi, S, Wu, S, Martek, I, Poshdar, M & Ghodrati, N 2021, ‘Blockchain and the “Internet of Things’’ for the construction industry: research trends and opportunities”’, *Automation in Construction*, vol. 132, Elsevier, p. 103942.
98. Elia, G, Margherita, A & Secundo, G 2020, ‘Project management canvas: a systems thinking framework to address project complexity’, *International Journal of Managing Projects in Business*, vol. 14, Emerald Group Holdings Ltd., no. 4, pp. 809–835.
99. Elster, J 2023, ‘Policy-Development and Deference to Moral Experts’, *Res Publica*, Springer Science and Business Media B.V., pp. 1–19, viewed 13 May 2023, <https://link.springer.com/article/10.1007/s11158-022-09577-w>.
100. ‘E-management: Tools and Techniques of Leaders in Non-Profit Organizations for Online Volunteer Work’ 2022, *Revista de Management Comparat Internațional*, vol. 23, EDITURA ASE, no. 3, pp. 356–366.
101. *Emotional Intelligence as an Important Asset for HR in Organizations: Attitudes and Working Variables - ProQuest* n.d., viewed 6 May 2023, <https://www.proquest.com/openview/d8d1e48a7ac3126d21c83439b9207173/1?pq-origsite=gscholar&cbl=4425142>.
102. Enshassi, MSA, Walbridge, S, West, JS & Haas, CT 2019, ‘Integrated Risk Management Framework for Tolerance-Based Mitigation Strategy Decision Support in Modular Construction Projects’, *Journal of Management in Engineering*, vol. 35, American Society of Civil Engineers, no. 4, p. 05019004, viewed 12 May 2023, <https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29ME.1943-5479.0000698>.
103. Ericson, BJ, Denny, P, Prather, J, Duran, R, Hellas, A, Leinonen, J, Miller, CS, Morrison, BB, Pearce, JL & Rodger, SH 2022, ‘Parsons Problems and Beyond: Systematic Literature Review and Empirical Study Designs’, *Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE*, Association for Computing Machinery, pp. 191–234, viewed 12 May 2023, <https://dl.acm.org/doi/10.1145/3571785.3574127>.
104. Essam, N, Khodeir, L & Fathy, F 2023, ‘Approaches for BIM-based multi-objective optimization in construction scheduling’, *Ain Shams Engineering Journal*, vol. 14, Elsevier, no. 6, p. 102114.
105. Esteki, M, Gandomani, TJ & Farsani, HK 2020, ‘A risk management framework for distributed scrum using PRINCE2 methodology’, *Bulletin of Electrical Engineering and Informatics*, vol. 9,

Institute of Advanced Engineering and Science, no. 3, pp. 1299–1310, viewed 5 May 2023, <https://beei.org/index.php/EEI/article/view/1905>.

106. *Evaluating the non-academic impact of academic research: design considerations: Journal of Higher Education Policy and Management: Vol 39, No 1* n.d., viewed 16 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/1360080X.2016.1254429>.
107. Evans, C, Kandiko Howson, C, Forsythe, A & Edwards, C 2020, 'What constitutes high quality higher education pedagogical research?', *https://doi.org/10.1080/02602938.2020.1790500*, vol. 46, Routledge, no. 4, pp. 525–546, viewed 13 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/02602938.2020.1790500>.
108. Fakhar Manesh, M, Pellegrini, MM, Marzi, G & Dabic, M 2021, 'Knowledge Management in the Fourth Industrial Revolution: Mapping the Literature and Scoping Future Avenues', *IEEE Transactions on Engineering Management*, vol. 68, Institute of Electrical and Electronics Engineers Inc., no. 1, pp. 289–300.
109. Fan, G, Xia, J, Song, J, Sun, H & Liang, D 2022, 'Research on application of ecohydrology to disaster prevention and mitigation in China: a review', *Water Supply*, vol. 22, IWA Publishing, no. 3, pp. 2946–2958, viewed 4 May 2023, <http://iwaponline.com/ws/article-pdf/22/3/2946/1023202/ws022032946.pdf>.
110. Fan, J, Meng, J, Ludescher, J, Chen, X, Ashkenazy, Y, Kurths, J, Havlin, S & Schellnhuber, HJ 2021, 'Statistical physics approaches to the complex Earth system', *Physics Reports*, vol. 896, North-Holland, pp. 1–84.
111. Flores, E, Xu, X & Lu, Y 2020, 'Human Capital 4.0: a workforce competence typology for Industry 4.0', *Journal of Manufacturing Technology Management*, vol. 31, Emerald Group Holdings Ltd., no. 4, pp. 687–703.
112. Fobiri, G, Musonda, I & Muleya, F 2022, 'Reality Capture in Construction Project Management: A Review of Opportunities and Challenges', *Buildings 2022, Vol. 12, Page 1381*, vol. 12, Multidisciplinary Digital Publishing Institute, no. 9, p. 1381, viewed 4 May 2023, <https://www.mdpi.com/2075-5309/12/9/1381/htm>.
113. Forcael, E, González, V, Orozco, F, Opazo, A, Suazo, Á & Aránguiz, P 2014, 'Application of Problem-Based Learning to Teaching the Critical Path Method', *Journal of Professional Issues in Engineering Education and Practice*, vol. 141, American Society of Civil Engineers, no. 3, p. 04014016, viewed 20 April 2023, <https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29EI.1943-5541.0000236>.
114. Galetsi, P, Katsaliaki, K & Kumar, S 2020, 'Big data analytics in health sector: Theoretical framework, techniques and prospects', *International Journal of Information Management*, vol. 50, Pergamon, pp. 206–216.
115. Gao, J & Xu, B 2021, 'Complex Systems, Emergence, and Multiscale Analysis: A Tutorial and Brief Survey', *Applied Sciences 2021, Vol. 11, Page 5736*, vol. 11, Multidisciplinary Digital Publishing Institute, no. 12, p. 5736, viewed 12 May 2023, <https://www.mdpi.com/2076-3417/11/12/5736/htm>.
116. Garcia, J, Villavicencio, G, Altimiras, F, Crawford, B, Soto, R, Minatogawa, V, Franco, M, Martínez-Muñoz, D & Yepes, V 2022, 'Machine learning techniques applied to construction: A hybrid bibliometric analysis of advances and future directions', *Automation in Construction*, vol. 142, Elsevier, p. 104532.
117. Garg, P & Girdhar, N 2021, 'A systematic review on spam filtering techniques based on natural language processing framework', *Proceedings of the Confluence 2021: 11th International Conference on Cloud Computing, Data Science and Engineering*, Institute of Electrical and Electronics Engineers Inc., pp. 30–35.
118. Garzón, J, Pavón, J & Baldiris, S 2019, 'Systematic review and meta-analysis of augmented reality in educational settings', *Virtual Reality*, vol. 23, Springer London, no. 4, pp. 447–459, viewed 13 May 2023, <https://link.springer.com/article/10.1007/s10055-019-00379-9>.
119. Van Gassel, FJM, Láscaris-Comneno, T & Maas, GJ 2014, 'The conditions for successful automated collaboration in construction', *Automation in Construction*, vol. 39, pp. 85–92.

120. Gemino, A, Horner Reich, B & Serrador, PM 2020, 'Agile, Traditional, and Hybrid Approaches to Project Success: Is Hybrid a Poor Second Choice?', *https://doi.org/10.1177/8756972820973082*, vol. 52, SAGE PublicationsSage CA: Los Angeles, CA, no. 2, pp. 161–175, viewed 5 May 2023, <https://journals.sagepub.com/doi/abs/10.1177/8756972820973082?journalCode=pmxa>.
121. Ghorbani, A & Professor, A 2023, 'A Review of Successful Construction Project Managers' Competencies and Leadership Profile', *Journal of Rehabilitation in Civil Engineering*, vol. 11, Semnan University, no. 1, pp. 76–95, viewed 20 April 2023, <https://civiljournal.semnan.ac.ir/article_6409.html>.
122. Giabbanelli, PJ & Tawfik, AA 2020, 'Reducing the Gap Between the Conceptual Models of Students and Experts Using Graph-Based Adaptive Instructional Systems', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12425 LNCS, Springer Science and Business Media Deutschland GmbH, pp. 538–556, viewed 12 May 2023, <https://link.springer.com/chapter/10.1007/978-3-030-60128-7_40>.
123. Goldsack, T, Zhang, Z, Lin, C & Scarton, C 2022, 'Making Science Simple: Corpora for the Lay Summarisation of Scientific Literature', viewed 13 May 2023, <https://arxiv.org/abs/2210.09932v1>.
124. González-Hernández, IJ, Granillo-Macías, R, Rondero-Guerrero, C & Simón-Marmolejo, I 2021, 'Marshall-Olkin distributions: a bibliometric study', *Scientometrics 2021 126:11*, vol. 126, Springer, no. 11, pp. 9005–9029, viewed 7 May 2023, <https://link.springer.com/article/10.1007/s11192-021-04156-x>.
125. Govindaras, B, Wern, TS, Kaur, S, Haslin, IA & Ramasamy, RK 2023, 'Sustainable Environment to Prevent Burnout and Attrition in Project Management', *Sustainability 2023, Vol. 15, Page 2364*, vol. 15, Multidisciplinary Digital Publishing Institute, no. 3, p. 2364, viewed 4 May 2023, <https://www.mdpi.com/2071-1050/15/3/2364/htm>.
126. Grieves, M 2022, 'Intelligent digital twins and the development and management of complex systems', *Digital Twin*, vol. 2, F1000 Research Ltd, p. 8.
127. Grimble, R & Wellard, K 1997, 'Stakeholder methodologies in natural resource management: a review of principles, contexts, experiences and opportunities', *Agricultural Systems*, vol. 55, Elsevier, no. 2, pp. 173–193.
128. Gruson, Damien, Helleputte, T, Rousseau, P & Gruson, David 2019, 'Data science, artificial intelligence, and machine learning: Opportunities for laboratory medicine and the value of positive regulation', *Clinical Biochemistry*, vol. 69, Elsevier, pp. 1–7.
129. Guanci, G & Bjork, C 2019, 'An introduction to project management', *Nursing Management*, vol. 50, Lippincott Williams and Wilkins, no. 10, pp. 20–26, viewed 4 May 2023, <https://journals.lww.com/nursingmanagement/Fulltext/2019/10000/An_introduction_to_project_management.5.aspx>.
130. Gunn, A & Mintrom, M 2016, 'Evaluating the non-academic impact of academic research: design considerations', *http://dx.doi.org/10.1080/1360080X.2016.1254429*, vol. 39, Routledge, no. 1, pp. 20–30, viewed 16 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/1360080X.2016.1254429>.
131. Gupta, R, Tanwar, S, Tyagi, S & Kumar, N 2020, 'Machine Learning Models for Secure Data Analytics: A taxonomy and threat model', *Computer Communications*, vol. 153, Elsevier, pp. 406–440.
132. Gusenbauer, M & Haddaway, NR 2020, 'Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources', *Research Synthesis Methods*, vol. 11, John Wiley & Sons, Ltd, no. 2, pp. 181–217, viewed 13 May 2023, <https://onlinelibrary.wiley.com/doi/full/10.1002/jrsm.1378>.
133. Habibi Rad, M, Mojtahedi, M & Ostwald, MJ 2021, 'Industry 4.0, Disaster Risk Management and Infrastructure Resilience: A Systematic Review and Bibliometric Analysis', *Buildings 2021, Vol. 11, Page 411*, vol. 11, Multidisciplinary Digital Publishing Institute, no. 9, p. 411, viewed 7 May 2023, <https://www.mdpi.com/2075-5309/11/9/411/htm>.
134. Hadjinicolaou, N, Kader, M & Abdallah, I 2021, 'Strategic Innovation, Foresight and the Deployment of Project Portfolio Management under Mid-Range Planning Conditions in Medium-Sized Firms',

Sustainability 2022, Vol. 14, Page 80, vol. 14, Multidisciplinary Digital Publishing Institute, no. 1, p. 80, viewed 4 May 2023, <https://www.mdpi.com/2071-1050/14/1/80/htm>.

135. Hamouche, S 2021, ‘Human resource management and the COVID-19 crisis: implications, challenges, opportunities, and future organizational directions’, viewed 6 May 2023, <https://doi.org/10.1017/jmo.2021.15>.
136. Hariri, RH, Fredericks, EM & Bowers, KM 2019, ‘Uncertainty in big data analytics: survey, opportunities, and challenges’, *Journal of Big Data*, vol. 6, SpringerOpen, no. 1, pp. 1–16, viewed 7 May 2023, <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0206-3>.
137. Hartmann, NN & Lussier, B 2020, ‘Managing the sales force through the unexpected exogenous COVID-19 crisis’, *Industrial Marketing Management*, vol. 88, Elsevier, pp. 101–111.
138. Harzing, AW 2019, ‘Two new kids on the block: How do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science?’, *Scientometrics*, vol. 120, Springer Netherlands, no. 1, pp. 341–349, viewed 13 May 2023, <https://link.springer.com/article/10.1007/s11192-019-03114-y>.
139. He, Q, Wang, T, Chan, APC, Li, H & Chen, Y 2019, ‘Identifying the gaps in project success research: A mixed bibliographic and bibliometric analysis’, *Engineering, Construction and Architectural Management*, vol. 26, Emerald Group Holdings Ltd., no. 8, pp. 1553–1573.
140. He, Y, Yang, S, Chan, CY, Chen, L & Wu, C 2021, ‘Visualization Analysis of Intelligent Vehicles Research Field Based on Mapping Knowledge Domain’, *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, Institute of Electrical and Electronics Engineers Inc., no. 9, pp. 5721–5736.
141. *Henry L. Gantt and Frederick Taylor: The Pioneers of Scientific Management - ProQuest* n.d., viewed 20 April 2023, <https://www.proquest.com/openview/89786e634732c160409b9dc7f1454249/1?pq-origsite=gscholar&cbl=27161>.
142. Herath, M & Perera, P 2023, ‘Overview of Generic Project Management Methodologies, Frameworks, and Standards’, pp. 9–40, viewed 20 April 2023, <https://www.worldscientific.com/doi/10.1142/9789811240584_0002>.
143. Herath, S & Chong, SC 2021, ‘Key Components and Critical Success Factors for Project Management Success: A Literature Review’, *Operations and Supply Chain Management: An International Journal*, vol. 14, Operations and Supply Chain Management Forum, no. 4, pp. 431–443, viewed 20 April 2023, <https://www.journal.oscm-forum.org/publication/article/key-components-and-critical-success-factors-for-project-management-success-a-literature-review>.
144. Herndl, CG 1993, ‘Teaching Discourse and Reproducing Culture: A Critique of Research and Pedagogy in Professional and Non-Academic Writing’, *College Composition and Communication*, vol. 44, National Council of Teachers of English, no. 3, p. 349.
145. Holmlund, M, Van Vaerenbergh, Y, Ciuchita, R, Ravald, A, Sarantopoulos, P, Ordenes, FV & Zaki, M 2020, ‘Customer experience management in the age of big data analytics: A strategic framework’, *Journal of Business Research*, vol. 116, Elsevier, pp. 356–365.
146. Holzinger, A, Dehmer, M, Emmert-Streib, F, Cucchiara, R, Augenstein, I, Ser, J Del, Samek, W, Jurisica, I & Díaz-Rodríguez, N 2022, ‘Information fusion as an integrative cross-cutting enabler to achieve robust, explainable, and trustworthy medical artificial intelligence’, *Information Fusion*, vol. 79, Elsevier, pp. 263–278.
147. Holzinger, A, Keiblinger, K, Holub, P, Zatloukal, K & Müller, H 2023, ‘AI for life: Trends in artificial intelligence for biotechnology’, *New Biotechnology*, vol. 74, Elsevier, pp. 16–24.
148. Hong, Z, Ward, L, Chard, K, Blaiszik, B & Foster, I 2021, ‘Challenges and Advances in Information Extraction from Scientific Literature: a Review’, *JOM*, vol. 73, Springer, no. 11, pp. 3383–3400, viewed 10 May 2023, <https://link.springer.com/article/10.1007/s11837-021-04902-9>.
149. Hossain, NUI, Dayarathna, VL, Nagahi, M & Jaradat, R 2020, ‘Systems Thinking: A Review and Bibliometric Analysis’, *Systems 2020, Vol. 8, Page 23*, vol. 8, Multidisciplinary Digital Publishing Institute, no. 3, p. 23, viewed 12 May 2023, <https://www.mdpi.com/2079-8954/8/3/23/htm>.
150. Hou, R, Kong, YQ, Cai, B & Liu, H 2020, ‘Unstructured big data analysis algorithm and simulation of Internet of Things based on machine learning’, *Neural Computing and Applications*, vol. 32,

Springer, no. 10, pp. 5399–5407, viewed 7 May 2023, <https://link.springer.com/article/10.1007/s00521-019-04682-z>.

151. *How the Great Pyramid Was Built - Craig B. Smith - Google Books* n.d., viewed 20 April 2023, <https://books.google.com/books?hl=en&lr=&id=ErYsEAAAQBAJ&oi=fnd&pg=PA7&dq=The+Great+Pyramid+of+Giza,+which+was+built+in+2570+BC+and+is+considered+a+remarkable+achievement+of+ancient+project+managers.+&ots=vDEjUoS6X4&sig=hIJjegLGX9tpzRokgwBTkWGi8Rw#v=onepage&q&f=false>.
152. Hoyer, WD, Kroschke, M, Schmitt, B, Kraume, K & Shankar, V 2022, 'Transforming the Customer Experience through New Technologies', *https://doi.org/10.1016/j.intmar.2020.04.001*, vol. 51, SAGE PublicationsSage CA: Los Angeles, CA, pp. 57–71, viewed 7 May 2023, <https://journals.sagepub.com/doi/abs/10.1016/j.intmar.2020.04.001?journalCode=jnma>.
153. Hu, M & Milner, D 2020, 'Visualizing the research of embodied energy and environmental impact research in the building and construction field: A bibliometric analysis', *Developments in the Built Environment*, vol. 3, Elsevier, p. 100010.
154. Huang, J, Wu, X, Ling, S, Li, X, Wu, Y, Peng, L & He, Z 2022, 'A bibliometric and content analysis of research trends on GIS-based landslide susceptibility from 2001 to 2020', *Environmental Science and Pollution Research 2022 29:58*, vol. 29, Springer, no. 58, pp. 86954–86993, viewed 7 May 2023, <https://link.springer.com/article/10.1007/s11356-022-23732-z>.
155. Ikudayisi, AE, Chan, APC, Darko, A & Adegun, OB 2022, 'Integrated design process of green building projects: A review towards assessment metrics and conceptual framework', *Journal of Building Engineering*, vol. 50, Elsevier, p. 104180.
156. Imran, R & Soomro, TR 2022, 'Mapping of Agile Processes into Project Management Knowledge Areas and Processes', *2022 International Conference on Business Analytics for Technology and Security, ICBATS 2022*, Institute of Electrical and Electronics Engineers Inc.
157. Irtaymah, HA, Shaari, H & Ahmed, N 2022, 'Agile Project Management Approaches: A Case Study with Respect to Their Application in Finance Technology Projects', *Journal of Pure & Applied Sciences*, vol. 21, no. 4, pp. 24–29, viewed 20 April 2023, <http://sebhau.edu.ly/journal/index.php/jopas/article/view/2118>.
158. Jafari, KG & Noorzai, E 2021, 'Selecting the Most Appropriate Project Manager to Improve the Performance of the Occupational Groups in Road Construction Projects in Warm Regions', *Journal of Construction Engineering and Management*, vol. 147, American Society of Civil Engineers, no. 10, p. 04021131, viewed 4 May 2023, <https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29CO.1943-7862.0002151>.
159. Jamshed, S & Majeed, N 2019, 'Relationship between team culture and team performance through lens of knowledge sharing and team emotional intelligence', *Journal of Knowledge Management*, vol. 23, Emerald Group Holdings Ltd., no. 1, pp. 90–109.
160. Javaid, M, Haleem, A, Singh, RP & Suman, R 2022, 'Enhancing smart farming through the applications of Agriculture 4.0 technologies', *International Journal of Intelligent Networks*, vol. 3, Elsevier, pp. 150–164.
161. Jayagopal, V & Basser, KK 1AD, 'Data Management and Big Data Analytics: Data Management in Digital Economy', *https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-6684-3662-2.ch078*, vol. 4, IGI Global, pp. 1614–1633, viewed 7 May 2023, <https://www.igi-global.com/chapter/data-management-and-big-data-analytics/291055>.
162. Jayantha, WM & Oladinrin, OT 2020, 'Bibliometric analysis of hedonic price model using CiteSpace', *International Journal of Housing Markets and Analysis*, vol. 13, Emerald Group Holdings Ltd., no. 2, pp. 357–371.
163. Johnson, JL, Adkins, D & Chauvin, S 2020, 'A Review of the Quality Indicators of Rigor in Qualitative Research', *American Journal of Pharmaceutical Education*, vol. 84, American Journal of Pharmaceutical Education, no. 1, pp. 138–146, viewed 13 May 2023, <https://www.ajpe.org/content/84/1/7120>.

164. Kalyan, KS, Rajasekharan, A & Sangeetha, S 2021, 'AMMUS : A Survey of Transformer-based Pretrained Models in Natural Language Processing', viewed 10 May 2023, <https://arxiv.org/abs/2108.05542v2>.
165. Kang, Y, Cai, Z, Tan, CW, Huang, Q & Liu, H 2020, 'Natural language processing (NLP) in management research: A literature review', *https://doi.org/10.1080/23270012.2020.1756939*, vol. 7, Taylor & Francis, no. 2, pp. 139–172, viewed 10 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/23270012.2020.1756939>.
166. Karhadkar, A, Laddha, A, Revanth, NM, Roy Choudhury, A & Panimozhi, K 2022, 'NLP Based Review Categorization: A Survey', *Proceedings - 2022 6th International Conference on Intelligent Computing and Control Systems, ICICCS 2022*, Institute of Electrical and Electronics Engineers Inc., pp. 1516–1522.
167. Kaushik, P & Datta, B 2020, 'Applicability of Agile Methodology in Technology Migration Projects: A Thematic Overview', *Asian Journal of Management*, vol. 11, A and V Publications, no. 1, p. 97.
168. Kelleher, AH 2004, 'An Investigation of the Expanding Role of the Critical Path Method by ENR'S Top 400 Contractors', Virginia Tech, viewed 20 April 2023, <https://vtechworks.lib.vt.edu/handle/10919/9889>.
169. Kemeç, A & Altınay, AT 2023, 'Sustainable Energy Research Trend: A Bibliometric Analysis Using VOSviewer, RStudio Bibliometrix, and CiteSpace Software Tools', *Sustainability 2023, Vol. 15, Page 3618*, vol. 15, Multidisciplinary Digital Publishing Institute, no. 4, p. 3618, viewed 7 May 2023, <https://www.mdpi.com/2071-1050/15/4/3618/htm>.
170. Kendzierskyj, S, Jahankhani, H, Jamal, A & Jimenez, JI 2019, 'The transparency of big data, data harvesting and digital Twins', *Advanced Sciences and Technologies for Security Applications*, Springer, pp. 139–148, viewed 7 May 2023, <https://link.springer.com/chapter/10.1007/978-3-030-11289-9_6>.
171. Kenley, R & Seppänen, O 2009, 'Location-based management of construction projects: Part of a new typology for project scheduling methodologies', *Proceedings - Winter Simulation Conference*, pp. 2563–2570.
172. Kermanshachi, S, Nipa, TJ & Dao, B 2021, 'Development of complexity management strategies for construction projects', *Journal of Engineering, Design and Technology*, vol. ahead-of-print, Emerald Group Holdings Ltd., no. ahead-of-print.
173. Khatib, M El, Almteiri, M & Qasemi, SA Al 2021, 'The Correlation between Emotional Intelligence and Project Management Success', *iBusiness*, vol. 13, Scientific Research Publishing, no. 01, pp. 18–29, viewed 5 May 2023, <//file.scirp.org/html/2-8601600_107331.htm>.
174. Khatter, A, Naughton, M, Dambha-Miller, H & Redmond, P 2021, 'Is rapid scientific publication also high quality? Bibliometric analysis of highly disseminated COVID-19 research papers', *Learned Publishing*, vol. 34, John Wiley & Sons, Ltd, no. 4, pp. 568–577, viewed 13 May 2023, <https://onlinelibrary.wiley.com/doi/full/10.1002/leap.1403>.
175. Khodabandelu, A & Park, JW 2021, 'Agent-based modeling and simulation in construction', *Automation in Construction*, vol. 131, Elsevier, p. 103882.
176. Kim, B 2020, 'Moving Forward with Digital Disruption: What Big Data, IoT, Synthetic Biology, AI, Blockchain, and Platform Businesses Mean to Libraries', *Technical Services Department Faculty Publications*, vol. 56, no. 2, viewed 7 May 2023, <https://digitalcommons.uri.edu/lib_ts_pubs/122>.
177. Kim, JM, Bae, J, Son, S, Son, K & Yum, SG 2021, 'Development of Model to Predict Natural Disaster-Induced Financial Losses for Construction Projects Using Deep Learning Techniques', *Sustainability 2021, Vol. 13, Page 5304*, vol. 13, Multidisciplinary Digital Publishing Institute, no. 9, p. 5304, viewed 7 May 2023, <https://www.mdpi.com/2071-1050/13/9/5304/htm>.
178. Kineber, AF, Othman, I, Oke, AE, Chileshe, N & Zayed, T 2023, 'Value management implementation barriers for sustainable building: a bibliometric analysis and partial least square structural equation modeling', *Construction Innovation*, vol. 23, Emerald Publishing, no. 1, pp. 38–73.
179. Kington, RS, Arnesen, S, Chou, W-YS, Curry, SJ, Lazer, D & Villarruel, AM 2021, 'Identifying Credible Sources of Health Information in Social Media: Principles and Attributes', *NAM*

Perspectives, vol. 2021, National Academy of Medicine, viewed 13 May 2023, </pmc/articles/PMC8486420/>.
180. Kircaburun, K, Harris, A, Calado, F & Griffiths, MD 2021, 'The Psychology of Mukbang Watching: A Scoping Review of the Academic and Non-academic Literature', *International Journal of Mental Health and Addiction*, vol. 19, Springer, no. 4, pp. 1190–1213, viewed 16 May 2023, <https://link.springer.com/article/10.1007/s11469-019-00211-0>.
181. Kononova, O, He, T, Huo, H, Trewartha, A, Olivetti, EA & Ceder, G 2021, 'Opportunities and challenges of text mining in materials research', *iScience*, vol. 24, Elsevier, no. 3, p. 102155.
182. Kretser, A, Murphy, D, Bertuzzi, S, Abraham, T, Allison, DB, Boor, KJ, Dwyer, J, Grantham, A, Harris, LJ, Hollander, R, Jacobs-Young, C, Rovito, S, Vafiadis, D, Woteki, C, Wyndham, J & Yada, R 2019, 'Scientific Integrity Principles and Best Practices: Recommendations from a Scientific Integrity Consortium', *Science and Engineering Ethics*, vol. 25, Springer Netherlands, no. 2, pp. 327–355, viewed 13 May 2023, <https://link.springer.com/article/10.1007/s11948-019-00094-3>.
183. Kumar, S & Markscheffel, B 2016, 'Bonded-communities in HantaVirus research: a research collaboration network (RCN) analysis', *Scientometrics*, vol. 109, Springer Netherlands, no. 1, pp. 533–550, viewed 13 November 2023, <https://link.springer.com/article/10.1007/s11192-016-1942-1>.
184. Lähnemann, D, Köster, J, Szczurek, E, McCarthy, DJ, Hicks, SC, Robinson, MD, Vallejos, CA, Campbell, KR, Beerenwinkel, N, Mahfouz, A, Pinello, L, Skums, P, Stamatakis, A, Attolini, CSO, Aparicio, S, Baaijens, J, Balvert, M, Barbanson, B de, Cappuccio, A, Corleone, G, Dutilh, BE, Florescu, M, Guryev, V, Holmer, R, Jahn, K, Lobo, TJ, Keizer, EM, Khatri, I, Kielbasa, SM, Korbel, JO, Kozlov, AM, Kuo, TH, Lelieveldt, BPF, Mandoiu, II, Marioni, JC, Marschall, T, Mölder, F, Niknejad, A, Raczkowski, L, Reinders, M, Ridder, J de, Saliba, AE, Somarakis, A, Stegle, O, Theis, FJ, Yang, H, Zelikovsky, A, McHardy, AC, Raphael, BJ, Shah, SP & Schönhuth, A 2020, 'Eleven grand challenges in single-cell data science', *Genome Biology 2020 21:1*, vol. 21, BioMed Central, no. 1, pp. 1–35, viewed 7 May 2023, <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-1926-6>.
185. Larson, L & DeChurch, LA 2020, 'Leading teams in the digital age: Four perspectives on technology and what they mean for leading teams', *The Leadership Quarterly*, vol. 31, JAI, no. 1, p. 101377.
186. Lee, M, Kwon, W & Back, KJ 2021, 'Artificial intelligence for hospitality big data analytics: developing a prediction model of restaurant review helpfulness for customer decision-making', *International Journal of Contemporary Hospitality Management*, vol. 33, Emerald Group Holdings Ltd., no. 6, pp. 2117–2136.
187. Lee, SM & Lee, DH 2020, '"Untact": a new customer service strategy in the digital age', *Service Business*, vol. 14, Springer, no. 1, pp. 1–22, viewed 7 May 2023, <https://link.springer.com/article/10.1007/s11628-019-00408-2>.
188. Lemenkova, P 2022, 'Handling Dataset with Geophysical and Geological Variables on the Bolivian Andes by the GMT Scripts', *Data 2022, Vol. 7, Page 74*, vol. 7, Multidisciplinary Digital Publishing Institute, no. 6, p. 74, viewed 7 May 2023, <https://www.mdpi.com/2306-5729/7/6/74/htm>.
189. Leong, J, Yee, KM, Baitsegi, O, Palanisamy, L & Ramasamy, RK 2023, 'Hybrid Project Management between Traditional Software Development Lifecycle and Agile Based Product Development for Future Sustainability', *Sustainability 2023, Vol. 15, Page 1121*, vol. 15, Multidisciplinary Digital Publishing Institute, no. 2, p. 1121, viewed 4 May 2023, <https://www.mdpi.com/2071-1050/15/2/1121/htm>.
190. Leyer, M & Schneider, S 2021, 'Decision augmentation and automation with artificial intelligence: Threat or opportunity for managers?', *Business Horizons*, vol. 64, Elsevier, no. 5, pp. 711–724.
191. Li, L, Lin, J, Ouyang, Y & Luo, X (Robert) 2022, 'Evaluating the impact of big data analytics usage on the decision-making quality of organizations', *Technological Forecasting and Social Change*, vol. 175, North-Holland, p. 121355.
192. Liang, X, Luo, L, Hu, S & Li, Y 2022, 'Mapping the knowledge frontiers and evolution of decision making based on agent-based modeling', *Knowledge-Based Systems*, vol. 250, Elsevier, p. 108982.
193. Liao, L, Yang, C & Quan, L 2023, 'Construction supply chain management: A systematic literature review and future development', *Journal of Cleaner Production*, vol. 382, Elsevier, p. 135230.

194. Liu, D, Wang, Y, Chen, J & Zhang, Y 2020, 'Intelligent compaction practice and development: a bibliometric analysis', *Engineering, Construction and Architectural Management*, vol. 27, Emerald Group Holdings Ltd., no. 5, pp. 1213–1232.
195. Liu, Y, Soroka, A, Han, L, Jian, J & Tang, M 2020, 'Cloud-based big data analytics for customer insight-driven design innovation in SMEs', *International Journal of Information Management*, vol. 51, Pergamon, p. 102034.
196. Locke, S, Bashall, A, Al-Adely, S, Moore, J, Wilson, A & Kitchen, GB 2021, 'Natural language processing in medicine: A review', *Trends in Anaesthesia and Critical Care*, vol. 38, Elsevier, pp. 4–9.
197. López-Pernas, S, Saqr, M & Apiola, M 2023, 'Scientometrics: A Concise Introduction and a Detailed Methodology for Mapping the Scientific Field of Computing Education Research', *Past, Present and Future of Computing Education Research*, Springer, Cham, Cham, pp. 79–99, viewed 7 May 2023, <https://link.springer.com/chapter/10.1007/978-3-031-25336-2_5>.
198. López-Úbeda, P, Martín-Noguerol, T, Aneiros-Fernández, J & Luna, A 2022, 'Natural Language Processing in Pathology: Current Trends and Future Insights', *The American Journal of Pathology*, vol. 192, Elsevier, no. 11, pp. 1486–1495.
199. Lu, J, Cairns, L & Smith, L 2020, 'Data science in the business environment: customer analytics case studies in SMEs', *Journal of Modelling in Management*, vol. 16, Emerald Group Holdings Ltd., no. 2, pp. 689–713.
200. Lu, M, Lam, HC & Dai, F 2008, 'Resource-constrained critical path analysis based on discrete event simulation and particle swarm optimization', *Automation in Construction*, vol. 17, Elsevier, no. 6, pp. 670–681.
201. Ma, X, Zhou, H & Li, Z 2021, 'On the resilience of modern power systems: A complex network perspective', *Renewable and Sustainable Energy Reviews*, vol. 152, Pergamon, p. 111646.
202. Maheshwari, S, Gautam, P & Jaggi, CK 2020, 'Role of Big Data Analytics in supply chain management: current trends and future perspectives', *https://doi.org/10.1080/00207543.2020.1793011*, vol. 59, Taylor & Francis, no. 6, pp. 1875–1900, viewed 7 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/00207543.2020.1793011>.
203. Malabagi, S, Kulkarni, VN, Gaitonde, VN, Satish, GJ & Kotturshettar, BB 2021, 'Product lifecycle management (PLM): A decision-making tool for project management', *AIP Conference Proceedings*, vol. 2358, American Institute of Physics Inc., no. 1, viewed 4 May 2023, <https://pubs.aip.org/aip/acp/article/2358/1/100013/1000699/Product-lifecycle-management-PLM-A-decision-making>.
204. Malla, V & Prasad, K V. 2022, 'Critical success factors and perceptions of stake holders to Lean-Agile adoption in construction projects (LACP)', *AIP Conference Proceedings*, vol. 2393, American Institute of Physics Inc., no. 1, viewed 5 May 2023, <https://pubs.aip.org/aip/acp/article/2393/1/020017/2821324/Critical-success-factors-and-perceptions-of-stake>.
205. Malyusz, L, Hajdu, M & Vattai, Z 2021, 'Comparison of different algorithms for time analysis for CPM schedule networks', *Automation in Construction*, vol. 127, Elsevier, p. 103697.
206. *Managing the Implementation of Business Intelligence Systems: A Critical Success Factors Framework: Business & Management Journal Article | IGI Global* n.d., viewed 7 May 2023, <https://www.igi-global.com/article/...business-intelligence-systems/2147>.
207. *Marketing Analytics: A Machine Learning Approach - Google Books* n.d., viewed 7 May 2023, <https://books.google.com/books?hl=en&lr=&id=lEuhEAAAQBAJ&oi=fnd&pg=PA303&dq=Unstructured+data+includes+text+documents,+images,+videos,+social+media+posts,+webpages+and+other+types+of+content+that+are+not+easily+quantifiable.+&ots=xLuJsryEUE&sig=dcDc0xB6-Hs7wy8tIfYVo-Jh4pE#v=onepage&q&f=false>.
208. Marle, F 2020, 'An assistance to project risk management based on complex systems theory and agile project management', *Complexity*, vol. 2020, Hindawi Limited.

209. Marnada, P, Raharjo, T, Hardian, B & Prasetyo, A 2022, ‘Agile project management challenge in handling scope and change: A systematic literature review’, *Procedia Computer Science*, vol. 197, Elsevier, pp. 290–300.
210. Maroufkhani, P, Tseng, ML, Iranmanesh, M, Ismail, WKW & Khalid, H 2020, ‘Big data analytics adoption: Determinants and performances among small to medium-sized enterprises’, *International Journal of Information Management*, vol. 54, Pergamon, p. 102190.
211. Martin, A 2023, ‘Introduction to an agile framework for the management of technology transfer projects’, *Procedia Computer Science*, vol. 219, Elsevier, pp. 1963–1968, viewed 20 April 2023, <https://linkinghub.elsevier.com/retrieve/pii/S1877050923005094>.
212. Martinez-Plumed, F, Contreras-Ochando, L, Ferri, C, Hernandez-Orallo, J, Kull, M, Lachiche, N, Ramirez-Quintana, MJ & Flach, P 2021, ‘CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories’, *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, IEEE Computer Society, no. 8, pp. 3048–3061.
213. Martín-Martín, A, Thelwall, M, Orduna-Malea, E & Delgado López-Cózar, E 2021, ‘Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations’ COCI: a multidisciplinary comparison of coverage via citations’, *Scientometrics*, vol. 126, Springer Science and Business Media B.V., no. 1, pp. 871–906, viewed 13 May 2023, <https://link.springer.com/article/10.1007/s11192-020-03690-4>.
214. Maruccia, Y, Solazzo, G, Del Vecchio, P & Passiante, G 2020, ‘Evidence from Network Analysis application to Innovation Systems and Quintuple Helix’, *Technological Forecasting and Social Change*, vol. 161, North-Holland, p. 120306.
215. Marvin, HJP, Kleter, GA, Frewer, LJ, Cope, S, Wentholt, MTA & Rowe, G 2009, ‘A working procedure for identifying emerging food safety issues at an early stage: Implications for European and international risk management practices’, *Food Control*, vol. 20, Elsevier, no. 4, pp. 345–356.
216. Mavi, RK, Gengatharen, D, Mavi, NK, Hughes, R, Campbell, A & Yates, R 2021, ‘Sustainability in Construction Projects: A Systematic Literature Review’, *Sustainability 2021, Vol. 13, Page 1932*, vol. 13, Multidisciplinary Digital Publishing Institute, no. 4, p. 1932, viewed 4 May 2023, <https://www.mdpi.com/2071-1050/13/4/1932/htm>.
217. Mazars, T & Francis, A 2020, ‘Chronographical spatiotemporal dynamic 4D planning’, *Automation in Construction*, vol. 112, Elsevier, p. 103076.
218. McGee, AA, Markarian, MD & Hamblen, JW 1962, ‘Business and management data processing I’, Association for Computing Machinery (ACM), pp. 56–57.
219. McIntyre, JC, Worsley, J, Corcoran, R, Harrison Woods, P & Bentall, RP 2018, ‘Academic and non-academic predictors of student psychological distress: the role of social identity and loneliness’, *https://doi.org/10.1080/09638237.2018.1437608*, vol. 27, Routledge, no. 3, pp. 230–239, viewed 16 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/09638237.2018.1437608>.
220. Medeiros, MM de & Maçada, ACG 2022, ‘Competitive advantage of data-driven analytical capabilities: the role of big data visualization and of organizational agility’, *Management Decision*, vol. 60, Emerald Group Holdings Ltd., no. 4, pp. 953–975.
221. Mendonça, S, Pereira, J & Ferreira, ME 2018, ‘Gatekeeping African studies: what does “editormetrics” indicate about journal governance?’, *Scientometrics*, vol. 117, Springer Netherlands, no. 3, pp. 1513–1534, viewed 16 May 2023, <https://link.springer.com/article/10.1007/s11192-018-2909-1>.
222. Merga, M & Mason, S 2020, ‘Early career researchers’ perceptions of the benefits and challenges of sharing research with academic and non-academic end-users’, *https://doi.org/10.1080/07294360.2020.1815662*, vol. 40, Routledge, no. 7, pp. 1482–1496, viewed 16 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/07294360.2020.1815662>.
223. Mergel, I, Ganapati, S & Whitford, AB 2021, ‘Agile: A New Way of Governing’, *Public Administration Review*, vol. 81, John Wiley & Sons, Ltd, no. 1, pp. 161–165, viewed 5 May 2023, <https://onlinelibrary.wiley.com/doi/full/10.1111/puar.13202>.

224. Micheni, E, MacHii, J & Murumba, J 2022, ‘Internet of Things, Big Data Analytics, and Deep Learning for Sustainable Precision Agriculture’, *2022 IST-Africa Conference, IST-Africa 2022*, Institute of Electrical and Electronics Engineers Inc.
225. Mikalef, P, Boura, M, Lekakos, G & Krogstie, J 2019, ‘Big Data Analytics Capabilities and Innovation: The Mediating Role of Dynamic Capabilities and Moderating Effect of the Environment’, *British Journal of Management*, vol. 30, John Wiley & Sons, Ltd, no. 2, pp. 272–298, viewed 7 May 2023, <https://onlinelibrary.wiley.com/doi/full/10.1111/1467-8551.12343>.
226. Mikalef, P, Krogstie, J, Pappas, IO & Pavlou, P 2020, ‘Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities’, *Information & Management*, vol. 57, North-Holland, no. 2, p. 103169.
227. Mishra, S & Tyagi, AK 2022, ‘The Role of Machine Learning Techniques in Internet of Things-Based Cloud Applications’, *Internet of Things*, Springer Science and Business Media Deutschland GmbH, pp. 105–135, viewed 7 May 2023, <https://link.springer.com/chapter/10.1007/978-3-030-87059-1_4>.
228. Monaghesh, E & Hajizadeh, A 2020, ‘The role of telehealth during COVID-19 outbreak: A systematic review based on current evidence’, *BMC Public Health*, vol. 20, BioMed Central, no. 1, pp. 1–9, viewed 13 May 2023, <https://link.springer.com/articles/10.1186/s12889-020-09301-4>.
229. Moradi, S, Kähkönen, K & Aaltonen, K 2020, ‘Comparison of research and industry views on project managers’ competencies’, *International Journal of Managing Projects in Business*, vol. 13, Emerald Group Holdings Ltd., no. 3, pp. 543–572.
230. Mordecai, Y, Fairbanks, J & Crawley, EF 2021, ‘Category-Theoretic Formulation of the Model-Based Systems Architecting Cognitive-Computational Cycle’, *Applied Sciences 2021, Vol. 11, Page 1945*, vol. 11, Multidisciplinary Digital Publishing Institute, no. 4, p. 1945, viewed 12 May 2023, <https://www.mdpi.com/2076-3417/11/4/1945/htm>.
231. Moselhi, O, Bardareh, H & Zhu, Z 2020, ‘Automated data acquisition in construction with remote sensing technologies’, *Applied Sciences (Switzerland)*, vol. 10, MDPI AG, no. 8.
232. Moskalenko, O, Muzychko, L, Hachak-Velychko, L, Dovzhuk, V & Blokhina, I 2022, ‘Application of International Scientometric Databases in the Professional Training of Future Teachers of Psychological and Pedagogical Disciplines: Capabilities of Web of science (WOS), Scopus, Google Scholar’, *IJCSNS International Journal of Computer Science and Network Security*, vol. 22, no. 5, p. 583, viewed 13 May 2023, <https://doi.org/10.22937/IJCSNS.2022.22.5.80>.
233. Motulsky, SL 2021, ‘Is Member Checking the Gold Standard of Quality in Qualitative Research?’, *Qualitative Psychology*, vol. 8, American Psychological Association, no. 3, pp. 389–406.
234. Muniswamaiah, M, Agerwala, T & Tappert, C 2019, ‘Big Data in Cloud Computing Review and Opportunities’, *International Journal of Computer Science and Information Technology*, vol. 11, Academy and Industry Research Collaboration Center (AIRCC), no. 4, pp. 43–57, viewed 7 May 2023, <http://arxiv.org/abs/1912.10821>.
235. Murphy, J, Dickerson, C, Thorpe, T, Goodier, C, Zahiroddiny, S & Pestana, A 2022, ‘Complexity identification in major infrastructure project information systems using graph theory’, *INCOSE International Symposium*, vol. 32, John Wiley & Sons, Ltd, no. S2, pp. 16–29, viewed 12 May 2023, <https://onlinelibrary.wiley.com/doi/full/10.1002/iis2.12892>.
236. Musawir, A ul, Abd-Karim, SB & Mohd-Danuri, MS 2020, ‘Project governance and its role in enabling organizational strategy implementation: A systematic literature review’, *International Journal of Project Management*, vol. 38, Pergamon, no. 1, pp. 1–16.
237. Mysirlaki, S & Paraskeva, F 2020, ‘Emotional intelligence and transformational leadership in virtual teams: lessons from MMOGs’, *Leadership and Organization Development Journal*, vol. 41, Emerald Group Holdings Ltd., no. 4, pp. 551–566.
238. Nabipour, M, Nayyeri, P, Jabani, H, Shahab, S & Mosavi, A 2020, ‘Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; A Comparative Analysis’, *IEEE Access*, vol. 8, Institute of Electrical and Electronics Engineers Inc., pp. 150199–150212.

239. Nabki, MW Al, Fidalgo, E, Alegre, E & De Paz, I 2017, ‘Classifying illegal activities on tor network based on web textual contents’, *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Conference*, vol. 1, Association for Computational Linguistics (ACL), pp. 35–43.
240. Nadikattu, RR 2020, ‘Research on Data Science, Data Analytics and Big Data’, *SSRN Electronic Journal*, Elsevier BV, viewed 7 May 2023, <https://papers.ssrn.com/abstract=3622844>.
241. Naeem, M, Jamal, T, Diaz-Martinez, J, Butt, SA, Montesano, N, Tariq, MI, De-la-Hoz-Franco, E & De-La-Hoz-Valdiris, E 2022, ‘Trends and Future Perspective Challenges in Big Data’, *Smart Innovation, Systems and Technologies*, vol. 253, Springer Science and Business Media Deutschland GmbH, pp. 309–325, viewed 7 May 2023, <https://link.springer.com/chapter/10.1007/978-981-16-5036-9_30>.
242. Najib, NH, Ismail, S, Che Amat, R, Durdyev, S, Konečná, Z, Chofreh, AG, Goni, FA, Subramaniam, C & Klemeš, JJ 2022, ‘Stakeholders’ Impact Factors of the COVID-19 Pandemic on Sustainable Mixed Development Projects: A Systematic Review and Meta-Analysis’, *Sustainability (Switzerland)*, vol. 14, MDPI, no. 16, p. 10418, viewed 4 May 2023, <https://www.mdpi.com/2071-1050/14/16/10418/htm>.
243. Nandwani, P & Verma, R 2021, ‘A review on sentiment analysis and emotion detection from text’, *Social Network Analysis and Mining*, vol. 11, Springer, no. 1, pp. 1–19, viewed 7 May 2023, <https://link.springer.com/article/10.1007/s13278-021-00776-6>.
244. Naseem, U, Razzak, I, Khan, SK & Prasad, M 2021, ‘A Comprehensive Survey on Word Representation Models: From Classical to State-of-the-Art Word Representation Language Models’, *Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, ACM PUB27 New York, NY , no. 5, viewed 10 May 2023, <https://dl.acm.org/doi/10.1145/3434237>.
245. Nedjwa, E, Bertrand, R & Sassi Boudemagh, S 2022, ‘Impacts of Industry 4.0 technologies on Lean management tools: a bibliometric analysis’, *International Journal on Interactive Design and Manufacturing*, vol. 16, Springer-Verlag Italia s.r.l., no. 1, pp. 135–150, viewed 7 May 2023, <https://link.springer.com/article/10.1007/s12008-021-00795-9>.
246. Ni, J, Lin, X & Shen, XS 2019, ‘Toward Edge-Assisted Internet of Things: From Security and Efficiency Perspectives’, *IEEE Network*, vol. 33, Institute of Electrical and Electronics Engineers Inc., no. 2, pp. 50–57.
247. Nozari, H, Fallah, M, Kazemipoor, H & Najafi, SE 2021, ‘Big data analysis of IoT-based supply chain management considering FMCG industries’, *Business Informatics*, vol. 15, National Research University, Higher School of Econoimics, no. 1, pp. 78–96.
248. Ofori, DF 2013, ‘Project Management Practices and Critical Success Factors-A Developing Country Perspective’, *International Journal of Business and Management*, vol. 8, no. 21.
249. Oliveira, A, Feyzi Behnagh, R, Ni, L, Mohsinah, AA, Burgess, KJ & Guo, L 2019, ‘Emerging technologies as pedagogical tools for teaching and learning science: A literature review’, *Human Behavior and Emerging Technologies*, vol. 1, John Wiley & Sons, Ltd, no. 2, pp. 149–160, viewed 12 May 2023, <https://onlinelibrary.wiley.com/doi/full/10.1002/hbe2.141>.
250. Olivetti, EA, Cole, JM, Kim, E, Kononova, O, Ceder, G, Han, TYJ & Hiszpanski, AM 2020, ‘Data-driven materials research enabled by natural language processing and information extraction’, *Applied Physics Reviews*, vol. 7, American Institute of Physics Inc., no. 4, p. 41317, viewed 10 May 2023, <https://pubs.aip.org/aip/apr/article/7/4/041317/832109/Data-driven-materials-research-enabled-by-natural>.
251. Olivieri, H, Seppänen, O, Alves, T da CL, Scala, NM, Schiavone, V, Liu, M & Granja, AD 2019, ‘Survey Comparing Critical Path Method, Last Planner System, and Location-Based Techniques’, *Journal of Construction Engineering and Management*, vol. 145, American Society of Civil Engineers, no. 12, p. 04019077, viewed 6 May 2023, <https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29CO.1943-7862.0001644>.
252. Olmos-Peñuela, J, Castro-Martínez, E & D’Este, P 2014, ‘Knowledge transfer activities in social sciences and humanities: Explaining the interactions of research groups with non-academic agents’, *Research Policy*, vol. 43, North-Holland, no. 4, pp. 696–706.

253. Ozkan, N, Bal, S, Erdogan, TG & Gok, MS 2022, ‘Scrum, Kanban or a Mix of Both? A Systematic Literature Review’, *Proceedings of the 17th Conference on Computer Science and Intelligence Systems, FedCSIS 2022*, Institute of Electrical and Electronics Engineers Inc., pp. 883–893.
254. Pais, S, Cordeiro, J & Jamil, ML 2022, ‘NLP-based platform as a service: a brief review’, *Journal of Big Data*, vol. 9, Springer Science and Business Media Deutschland GmbH, no. 1, pp. 1–26, viewed 10 May 2023, <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00603-5>.
255. Pan, X, Mateen Khan, A, Eldin, SM, Aslam, F, Kashif Ur Rehman, S & Jameel, M 2023, ‘BIM adoption in sustainability, energy modelling and implementing using ISO 19650: A review’, *Ain Shams Engineering Journal*, Elsevier, p. 102252.
256. Pan, Y & Zhang, L 2021, ‘A BIM-data mining integrated digital twin framework for advanced project management’, *Automation in Construction*, vol. 124, Elsevier, p. 103564.
257. Papadakis, E & Tsironis, L 2020, ‘Towards a hybrid project management framework: A systematic literature review on traditional, agile and hybrid techniques’, *The Journal of Modern Project Management*, vol. 8, no. 2, viewed 4 May 2023, <https://journalmodernpm.com/manuscript/index.php/jmpm/article/view/JMPM02410>.
258. Pascarella, ET & Terenzlni, PT 2014, ‘Informal Interaction with Faculty and Freshman Ratings of Academic and Non-Academic Experience of College 1’, *http://dx.doi.org/10.1080/00220671.1976.10884944*, vol. 70, Routledge, no. 1, pp. 35–41, viewed 16 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/00220671.1976.10884944>.
259. Paul, JA & Zhang, M 2021, ‘Decision support model for cybersecurity risk planning: A two-stage stochastic programming framework featuring firms, government, and attacker’, *European Journal of Operational Research*, vol. 291, North-Holland, no. 1, pp. 349–364.
260. Paul, K, Chatterjee, SS, Pai, P, Varshney, A, Juikar, S, Prasad, V, Bhadra, B & Dasgupta, S 2022, ‘Viable smart sensors and their application in data driven agriculture’, *Computers and Electronics in Agriculture*, vol. 198, Elsevier, p. 107096.
261. Payne-Sturges, DC, Cory-Slechta, DA, Puett, RC, Thomas, SB, Hammond, R & Hovmand, PS 2021, ‘Defining and Intervening on Cumulative Environmental Neurodevelopmental Risks: Introducing a Complex Systems Approach’, *Environmental Health Perspectives*, vol. 129, Environmental Health Perspectives , no. 3, viewed 12 May 2023, <https://ehp.niehs.nih.gov/doi/10.1289/EHP7333>.
262. Pei, Z, Chen, S, Ding, L, Liu, J, Cui, X, Li, F & Qiu, F 2022, ‘Current perspectives and trend of nanomedicine in cancer: A review and bibliometric analysis’, *Journal of Controlled Release*, vol. 352, Elsevier, pp. 211–241.
263. Posamentier, HW, Paumard, V & Lang, SC 2022, ‘Principles of seismic stratigraphy and seismic geomorphology I: Extracting geologic insights from seismic data’, *Earth-Science Reviews*, vol. 228, Elsevier, p. 103963.
264. Pouyanfar, S, Sadiq, S, Yan, Y, Tian, H, Tao, Y, Reyes, MP, Shyu, ML, Chen, SC & Iyengar, SS 2018, ‘A survey on deep learning: Algorithms, techniques, and applications’, *ACM Computing Surveys*, vol. 51, Association for Computing Machinery, no. 5.
265. Pranckutė, R 2021, ‘Web of Science (WoS) and Scopus: The Titans of Bibliographic Information in Today’s Academic World’, *Publications 2021, Vol. 9, Page 12*, vol. 9, Multidisciplinary Digital Publishing Institute, no. 1, p. 12, viewed 13 May 2023, <https://www.mdpi.com/2304-6775/9/1/12/htm>.
266. Prebanić, KR & Vukomanović, M 2021, ‘Realizing the Need for Digital Transformation of Stakeholder Management: A Systematic Review in the Construction Industry’, *Sustainability 2021, Vol. 13, Page 12690*, vol. 13, Multidisciplinary Digital Publishing Institute, no. 22, p. 12690, viewed 4 May 2023, <https://www.mdpi.com/2071-1050/13/22/12690/htm>.
267. Priyadarshini, I & Puri, V 2021, ‘A convolutional neural network (CNN) based ensemble model for exoplanet detection’, *Earth Science Informatics*, vol. 14, Springer Science and Business Media Deutschland GmbH, no. 2, pp. 735–747, viewed 7 May 2023, <https://link.springer.com/article/10.1007/s12145-021-00579-5>.
268. *Project Management Methodologies: Selecting, Implementing, and Supporting ... - Jason Charvat - Google Books* n.d., viewed 20 April 2023,

<https://books.google.fr/books?hl=en&lr=&id=Pd5AnH3ZA1AC&oi=fnd&pg=PR9&dq=The+project+managers+of+the+time+must+have+had+exceptional+skills+in+logistics,+resource+management,+and+leadership+to+oversee+such+a+complex+project+successfully.+&ots=Q9ESHoDfOC&sig=WuPle1UGDyq1YfRC8TRKxv5rU7U&redir_esc=y#v=onepage&q&f=false>.

269. Punjani, KK, Mahadevan, K, Gunasekaran, A, Kumar, VVR & Joshi, S 2022, 'Cloud computing in agriculture: a bibliometric and network visualization analysis', *Quality and Quantity*, Springer Science and Business Media B.V., pp. 1–35, viewed 7 May 2023, <https://link.springer.com/article/10.1007/s11135-022-01535-1>.
270. Qi, B, Razkenari, M, Costin, A, Kibert, C & Fu, M 2021, 'A systematic review of emerging technologies in industrialized construction', *Journal of Building Engineering*, vol. 39, Elsevier, p. 102265.
271. Radanliev, P, De Roure, D, Burnap, P & Santos, O 2021, 'Epistemological Equation for Analysing Uncontrollable States in Complex Systems: Quantifying Cyber Risks from the Internet of Things', *The Review of Socionetwork Strategies 2021 15:2*, vol. 15, Springer, no. 2, pp. 381–411, viewed 12 May 2023, <https://link.springer.com/article/10.1007/s12626-021-00086-5>.
272. Radini, S, González-Camejo, J, Andreola, C, Eusebi, AL & Fatone, F 2023, 'Risk management and digitalisation to overcome barriers for safe reuse of urban wastewater for irrigation – A review based on European practice', *Journal of Water Process Engineering*, vol. 53, Elsevier, p. 103690.
273. Rai, A & Borah, S 2021, 'Study of various methods for tokenization', *Lecture Notes in Networks and Systems*, vol. 137, Springer Science and Business Media Deutschland GmbH, pp. 193–200, viewed 10 May 2023, <https://link.springer.com/chapter/10.1007/978-981-15-6198-6_18>.
274. Raisch, S & Krakowski, S 2021, 'Artificial Intelligence and Management: The Automation–Augmentation Paradox', *https://doi.org/10.5465/amr.2018.0072*, vol. 46, Academy of ManagementBriarcliff Manor, NY, no. 1, pp. 192–210, viewed 5 May 2023, <https://journals.aom.org/doi/abs/10.5465/amr.2018.0072>.
275. Rampton, B, Maybin, J & Roberts, C 2015, 'Theory and Method in Linguistic Ethnography', *Linguistic Ethnography*, Palgrave Macmillan, London, pp. 14–50, viewed 16 May 2023, <https://link.springer.com/chapter/10.1057/9781137035035_2>.
276. Rehman, A, Naz, S & Razzak, I 2021, 'Leveraging big data analytics in healthcare enhancement: trends, challenges and opportunities', *Multimedia Systems 2021 28:4*, vol. 28, Springer, no. 4, pp. 1339–1371, viewed 7 May 2023, <https://link.springer.com/article/10.1007/s00530-020-00736-8>.
277. Ren, R, Hu, W, Dong, J, Sun, B, Chen, Y & Chen, Z 2019, 'A Systematic Literature Review of Green and Sustainable Logistics: Bibliometric Analysis, Research Trend and Knowledge Taxonomy', *International Journal of Environmental Research and Public Health 2020, Vol. 17, Page 261*, vol. 17, Multidisciplinary Digital Publishing Institute, no. 1, p. 261, viewed 7 May 2023, <https://www.mdpi.com/1660-4601/17/1/261/htm>.
278. Reponen, E, Rundall, TG, Shortell, SM, Blodgett, JC, Juarez, A, Jokela, R, Mäkijärvi, M & Torkki, P 2021, 'Benchmarking outcomes on multiple contextual levels in lean healthcare: a systematic review, development of a conceptual framework, and a research agenda', *BMC Health Services Research*, vol. 21, BioMed Central Ltd, no. 1, pp. 1–18, viewed 6 May 2023, <https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-021-06160-6>.
279. Rezvani, SMHS, Falcão, MJ, Komljenovic, D & de Almeida, NM 2023, 'A Systematic Literature Review on Urban Resilience Enabled with Asset and Disaster Risk Management Approaches and GIS-Based Decision Support Tools', *Applied Sciences 2023, Vol. 13, Page 2223*, vol. 13, Multidisciplinary Digital Publishing Institute, no. 4, p. 2223, viewed 4 May 2023, <https://www.mdpi.com/2076-3417/13/4/2223/htm>.
280. Ridzuan, F & Wan Zainon, WMN 2019, 'A Review on Data Cleansing Methods for Big Data', *Procedia Computer Science*, vol. 161, Elsevier, pp. 731–738.
281. Rodrigues, JFC, Ferreira, FAF, Pereira, LF, Carayannis, EG & Ferreira, JJM 2022, 'Banking Digitalization: (Re)Thinking Strategies and Trends Using Problem Structuring Methods', *IEEE Transactions on Engineering Management*, vol. 69, Institute of Electrical and Electronics Engineers Inc., no. 4, pp. 1517–1531.

282. Romzi, NA & Ing, DS 2022, ‘Key Constraint in Construction Industry: A Review’, *CONSTRUCTION*, vol. 2, Universiti Malaysia Pahang Publishing, no. 2, pp. 01–06, viewed 4 May 2023, <https://journal.ump.edu.my/construction/article/view/7779>.
283. Rose, J & Johnson, CW 2020, ‘Contextualizing reliability and validity in qualitative research: toward more rigorous and trustworthy qualitative social science in leisure research’, *https://doi.org/10.1080/00222216.2020.1722042*, vol. 51, Routledge, no. 4, pp. 432–451, viewed 13 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/00222216.2020.1722042>.
284. Roxas, CLC, Bautista, CR, Cruz, OG Dela, Cruz, RLC Dela, Pedro, JPQ De, Dungca, JR, Lejano, BA & Ongpeng, JMC 2023, ‘Design for Manufacturing and Assembly (DfMA) and Design for Deconstruction (DfD) in the Construction Industry: Challenges, Trends and Developments’, *Buildings 2023, Vol. 13, Page 1164*, vol. 13, Multidisciplinary Digital Publishing Institute, no. 5, p. 1164, viewed 12 May 2023, <https://www.mdpi.com/2075-5309/13/5/1164/htm>.
285. Rusthollkarhu, S, Toukola, S, Aarikka-Stenroos, L & Mahlamäki, T 2022, ‘Managing B2B customer journeys in digital era: Four management activities with artificial intelligence-empowered tools’, *Industrial Marketing Management*, vol. 104, Elsevier, pp. 241–257.
286. Safapour, E, Kermanshachi, S & Jafari, A 2021, ‘Investigating implementation level of phase-based strategies in transportation infrastructure projects’, *https://doi.org/10.1139/cjce-2021-0026*, vol. 49, Canadian Science Publishing 1840 Woodward Drive, Suite 1, Ottawa, ON K2C 0P7 , no. 5, pp. 792–801, viewed 4 May 2023, <https://cdnsciencepub.com/doi/10.1139/cjce-2021-0026>.
287. Santos, P de O & de Carvalho, MM 2022, ‘Exploring the challenges and benefits for scaling agile project management to large projects: a review’, *Requirements Engineering*, vol. 27, Springer Science and Business Media Deutschland GmbH, no. 1, pp. 117–134, viewed 4 May 2023, <https://link.springer.com/article/10.1007/s00766-021-00363-3>.
288. Santoso, B, Hikmawan, T & Imaniyati, N 2022, ‘Management Information Systems: Bibliometric Analysis and Its Effect on Decision Making’, *Indonesian Journal of Science and Technology*, vol. 7, Universitas Pendidikan Indonesia, no. 3, pp. 583–602, viewed 7 May 2023, <https://ejournal.upi.edu/index.php/ijost/article/view/56368>.
289. Sarker, IH 2021a, ‘Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective’, *SN Computer Science*, vol. 2, Springer, no. 5, pp. 1–22, viewed 7 May 2023, <https://link.springer.com/article/10.1007/s42979-021-00765-8>.
290. Sarker, IH 2021b, ‘Machine Learning: Algorithms, Real-World Applications and Research Directions’, *SN Computer Science*, vol. 2, Springer, no. 3, pp. 1–21, viewed 7 May 2023, <https://link.springer.com/article/10.1007/s42979-021-00592-x>.
291. Sarker, IH 2022, ‘Smart City Data Science: Towards data-driven smart cities with open research issues’, *Internet of Things*, vol. 19, Elsevier, p. 100528.
292. Sarker, IH, Hoque, MM, Uddin, MK & Alsanoosy, T 2021, ‘Mobile Data Science and Intelligent Apps: Concepts, AI-Based Modeling and Research Directions’, *Mobile Networks and Applications*, vol. 26, Springer, no. 1, pp. 285–303, viewed 7 May 2023, <https://link.springer.com/article/10.1007/s11036-020-01650-z>.
293. Sarker, IH, Kayes, ASM, Badsha, S, Alqahtani, H, Watters, P & Ng, A 2020, ‘Cybersecurity data science: an overview from machine learning perspective’, *Journal of Big Data*, vol. 7, Springer, no. 1, pp. 1–29, viewed 7 May 2023, <https://link.springer.com/articles/10.1186/s40537-020-00318-5>.
294. Saura, JR 2021, ‘Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics’, *Journal of Innovation & Knowledge*, vol. 6, Elsevier, no. 2, pp. 92–102.
295. Savio, R 2022, ‘Between Success and Failure: Integrating Agile Project Management Methods with PRINCE2 Framework for the Enhanced Management of Imperfect Projects’, *Eximia*, vol. 5, no. 1, pp. 24–32, viewed 20 April 2023, <https://www.eximiajournal.pluscommunication.eu/index.php/eximia/article/view/127>.
296. Sbiti, M, Beddiar, K, Beladjine, D, Perrault, R & Mazari, B 2021, ‘Toward BIM and LPS Data Integration for Lean Site Project Management: A State-of-the-Art Review and Recommendations’, *Buildings 2021, Vol. 11, Page 196*, vol. 11, Multidisciplinary Digital Publishing Institute, no. 5, p. 196, viewed 4 May 2023, <https://www.mdpi.com/2075-5309/11/5/196/htm>.

297. Scala, NM, Schiavone, V, Olivieri, H, Seppänen, O, Alves, T da CL, Liu, M & Granja, AD 2022, ‘Comparative Analysis of Planning with the Critical Path Method, Last Planner System, and Location-Based Techniques in Brazil, Finland, and the United States’, *https://doi.org/10.1080/10429247.2022.2069981*, Taylor & Francis, viewed 6 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/10429247.2022.2069981>.
298. Scharrer, L, Bromme, R & Stadtler, M 2021, ‘Information Easiness Affects Non-experts’ Evaluation of Scientific Claims About Which They Hold Prior Beliefs’, *Frontiers in Psychology*, vol. 12, Frontiers Media S.A., p. 3757.
299. Seaton, K, Loch, B & Lugosi, E 2022, ‘Takeaways from teaching through a global pandemic – practical examples of lasting value in tertiary mathematics education’, *https://doi.org/10.1080/0020739X.2022.2008551*, vol. 53, Taylor & Francis, no. 3, pp. 559–565, viewed 13 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/0020739X.2022.2008551>.
300. Senger, K, Betlem, P, Grundväg, SA, Horota, RK, Buckley, SJ, Smyrak-Sikora, A, Jochmann, MM, Birchall, T, Janocha, J, Ogata, K, Kuckero, L, Johannessen, RM, Lecomte, I, Cohen, SM & Olaussen, S 2021, ‘Teaching with digital geology in the high Arctic: Opportunities and challenges’, *Geoscience Communication*, vol. 4, Copernicus GmbH, no. 3, pp. 399–420.
301. Sergiou, C, Lestas, M, Antoniou, P, Liaskos, C & Pitsillides, A 2020, ‘Complex Systems: A Communication Networks Perspective towards 6G’, *IEEE Access*, vol. 8, Institute of Electrical and Electronics Engineers Inc., pp. 89007–89030.
302. Seyedan, M & Mafakheri, F 2020, ‘Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities’, *Journal of Big Data 2020 7:1*, vol. 7, SpringerOpen, no. 1, pp. 1–22, viewed 7 May 2023, <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00329-2>.
303. Seyhan, AA 2019, ‘Lost in translation: the valley of death across preclinical and clinical divide – identification of problems and overcoming obstacles’, *Translational Medicine Communications 2019 4:1*, vol. 4, BioMed Central, no. 1, pp. 1–19, viewed 7 May 2023, <https://link.springer.com/articles/10.1186/s41231-019-0050-7>.
304. Seymour, T & Hussein, S 2014, ‘The History Of Project Management’, *International Journal of Management & Information Systems (IJMIS)*, vol. 18, Clute Institute, no. 4, pp. 233–240, viewed 20 April 2023, <https://www.clutejournals.com/index.php/IJMIS/article/view/8820>.
305. Shahzad, K, Khan, SA, Ahmad, S & Iqbal, A 2022, ‘A Scoping Review of the Relationship of Big Data Analytics with Context-Based Fake News Detection on Digital Media in Data Age’, *Sustainability 2022, Vol. 14, Page 14365*, vol. 14, Multidisciplinary Digital Publishing Institute, no. 21, p. 14365, viewed 7 May 2023, <https://www.mdpi.com/2071-1050/14/21/14365/htm>.
306. Shaikh, I & Randhawa, K 2022, ‘Managing the risks and motivations of technology managers in open innovation: Bringing stakeholder-centric corporate governance into focus’, *Technovation*, vol. 114, Elsevier, p. 102437.
307. Sharifi, A 2021, ‘Urban sustainability assessment: An overview and bibliometric analysis’, *Ecological Indicators*, vol. 121, Elsevier, p. 107102.
308. Sharmila, Kumar, D, Kumar, P & Ashok, A 2020, ‘Introduction to multimedia big data computing for IoT’, *Intelligent Systems Reference Library*, vol. 163, Springer Science and Business Media Deutschland GmbH, pp. 3–36, viewed 7 May 2023, <https://link.springer.com/chapter/10.1007/978-981-13-8759-3_1>.
309. Shen, C, Wei, M & Sheng, Y 2021, ‘A bibliometric analysis of food safety governance research from 1999 to 2019’, *Food Science & Nutrition*, vol. 9, John Wiley & Sons, Ltd, no. 4, pp. 2316–2334, viewed 7 May 2023, <https://onlinelibrary.wiley.com/doi/full/10.1002/fsn3.2220>.
310. Sheng, J, Amankwah-Amoah, J, Khan, Z & Wang, X 2021, ‘COVID-19 Pandemic in the New Era of Big Data Analytics: Methodological Innovations and Future Research Directions’, *British Journal of Management*, vol. 32, John Wiley & Sons, Ltd, no. 4, pp. 1164–1183, viewed 7 May 2023, <https://onlinelibrary.wiley.com/doi/full/10.1111/1467-8551.12441>.

311. Shet, S V. & Pereira, V 2021, ‘Proposed managerial competencies for Industry 4.0 – Implications for social sustainability’, *Technological Forecasting and Social Change*, vol. 173, North-Holland, p. 121080.
312. Shetty, GS & Raghu, N 2022, ‘The Classification of Satellite Galaxies Based on Convolution Neural Networks Machine Learning Algorithm’, *2022 3rd International Conference for Emerging Technology, INCET 2022*, Institute of Electrical and Electronics Engineers Inc.
313. Singh, NP & Singh, S 2019, ‘Building supply chain risk resilience: Role of big data analytics in supply chain disruption mitigation’, *Benchmarking*, vol. 26, Emerald Group Holdings Ltd., no. 7, pp. 2318–2342.
314. Singh, VK, Singh, P, Karmakar, M, Leta, J & Mayr, P 2021, ‘The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis’, *Scientometrics*, vol. 126, Springer Science and Business Media B.V., no. 6, pp. 5113–5142, viewed 13 May 2023, <https://link.springer.com/article/10.1007/s11192-021-03948-5>.
315. Soegoto, H, Soegoto, ES, Luckyardi, S & Rafdhi, AA 2022, ‘A Bibliometric Analysis of Management Bioenergy Research Using Vosviewer Application’, *Indonesian Journal of Science and Technology*, vol. 7, Universitas Pendidikan Indonesia, no. 1, pp. 89–104, viewed 7 May 2023, <https://ejournal.upi.edu/index.php/ijost/article/view/43328>.
316. Somarathna, KUS 2020, ‘An agent-based approach for modeling and simulation of human resource management as a complex system: Management strategy evaluation’, *Simulation Modelling Practice and Theory*, vol. 104, Elsevier, p. 102118.
317. Stead, M, Coulton, P, Pilling, F, Gradinar, A, Pilling, M & Forrester, I 2022, ‘More-than-Human-Data Interaction’, *ACM International Conference Proceeding Series*, Association for Computing Machinery, pp. 62–74, viewed 7 May 2023, <https://dl.acm.org/doi/10.1145/3569219.3569344>.
318. Stephenson, PJ 2019, ‘The Holy Grail of biodiversity conservation management: Monitoring impact in projects and project portfolios’, *Perspectives in Ecology and Conservation*, vol. 17, Elsevier, no. 4, pp. 182–192.
319. Stoddard, MM, Gillis, B & Cohn, P 2019, ‘Agile Project Management in Libraries: Creating Collaborative, Resilient, Responsive Organizations’, *https://doi.org/10.1080/01930826.2019.1616971*, vol. 59, Routledge, no. 5, pp. 492–511, viewed 5 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/01930826.2019.1616971>.
320. Stone, C & Stone, C 2023, ‘Challenges and opportunities of completing successful projects using Earned Value Management’, *Open Journal of Business and Management*, vol. 11, Scientific Research Publishing, no. 2, pp. 464–493, viewed 4 May 2023, <http://www.scirp.org/journal/PaperInformation.aspx?PaperID=123579>.
321. Stylos, N, Zwiegelaar, J & Buhalis, D 2021, ‘Big data empowered agility for dynamic, volatile, and time-sensitive service industries: the case of tourism sector’, *International Journal of Contemporary Hospitality Management*, vol. 33, Emerald Group Holdings Ltd., no. 3, pp. 1015–1036.
322. Subrahmanya, SVG, Shetty, DK, Patil, V, Hameed, BMZ, Paul, R, Smriti, K, Naik, N & Somani, BK 2022, ‘The role of data science in healthcare advancements: applications, benefits, and future prospects’, *Irish Journal of Medical Science*, vol. 191, Springer Science and Business Media Deutschland GmbH, no. 4, pp. 1473–1483, viewed 7 May 2023, <https://link.springer.com/article/10.1007/s11845-021-02730-z>.
323. Suleman, Q, Hussain, I, Syed, MA, Parveen, R, Lodhi, IS & Mahmood, Z 2019, ‘Association between emotional intelligence and academic success among undergraduates: A cross-sectional study in KUST, Pakistan’, *PLOS ONE*, vol. 14, Public Library of Science, no. 7, p. e0219468, viewed 6 May 2023, <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0219468>.
324. Takagi, N & Varajão, J 2021, ‘ISO 21500 and success management: an integrated model for project management’, *International Journal of Quality & Reliability Management*, vol. 39, Emerald Publishing Limited, no. 2, pp. 408–427.
325. Tanglay, O, Dadario, NB, Chong, EHN, Tang, SJ, Young, IM & Sughrue, ME 2023, ‘Graph Theory Measures and Their Application to Neurosurgical Eloquence’, *Cancers 2023, Vol. 15, Page 556*, vol.

15, Multidisciplinary Digital Publishing Institute, no. 2, p. 556, viewed 12 May 2023, <https://www.mdpi.com/2072-6694/15/2/556/htm>.

326. Tariq, S, Hussein, M, Wang, RD & Zayed, T 2022, 'Trends and developments of on-site crane layout planning 1983–2020: bibliometric, scientometric and qualitative analyses', *Construction Innovation*, vol. 22, Emerald Publishing, no. 4, pp. 1011–1035.
327. Tavan, F & Hosseini, M 2016, 'Comparison and analysis of PMBOK 2013 and ISO 21500', *Journal of Project Management*, Growing Science, no. 10, pp. 27–34.
328. Teng, S & Khong, KW 2021, 'Examining actual consumer usage of E-wallet: A case study of big data analytics', *Computers in Human Behavior*, vol. 121, Pergamon, p. 106778.
329. *The AMA Handbook of Project Management - Paul C. Dinsmore, Jeannette Cabanis-Brewin - Google Books* n.d., viewed 20 April 2023, <https://books.google.fr/books?hl=en&lr=&id=qt6CAwAAQBAJ&oi=fnd&pg=PA13&dq=.+Two+of+the+most+widely+recognized+project+management+standards+are+the+Project+Management+Body+of+Knowledge+(PMBoK)+and+ISO+21500&ots=AERpU8qJN4&sig=hdNJFOrtlqB4QqxTPQbaG8dLjUA&redir_esc=y#v=onepage&q=.%20Two%20of%20the%20most%20widely%20recognized%20project%20management%20standards%20are%20the%20Project%20Management%20Body%20of%20Knowledge%20(PMBoK)%20and%20ISO%2021500&f=false>.
330. Timinger, H, Schmidtner, M & Reiche, F 2022, 'A Framework for the Construction and Tailoring of Engineering Development Process Models', *IEEE Transactions on Engineering Management*, Institute of Electrical and Electronics Engineers Inc.
331. Tuval, E & Isaac, S 2022, 'Online Planning and Management of Design Coordination Tasks with BIM: Challenges and Opportunities', *Journal of Management in Engineering*, vol. 38, American Society of Civil Engineers, no. 3, p. 05022003, viewed 4 May 2023, <https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29ME.1943-5479.0001027>.
332. Uzun, K 2022, 'Natural Language Processing Tools for Second Language Writing Research', *The TESOL Encyclopedia of English Language Teaching*, Wiley, pp. 1–6.
333. Valeri, M & Baggio, R 2021, 'Social network analysis: organizational implications in tourism management', *International Journal of Organizational Analysis*, vol. 29, Emerald Group Holdings Ltd., no. 2, pp. 342–353.
334. Varajão, J, Colomo-Palacios, R & Silva, H 2017, 'ISO 21500:2012 and PMBoK 5 processes in information systems project management', *Computer Standards & Interfaces*, vol. 50, North-Holland, pp. 216–222.
335. Vasconcelos, RN, Costa, DP, Duverger, SG, Lobão, JSB, Cambuí, ECB, Lentini, CAD, Lima, ATC, Schirmbeck, J, Mendes, DT, Rocha, WJSF & Porsani, MJ 2023, 'Bibliometric analysis of surface water detection and mapping using remote sensing in South America', *Scientometrics*, vol. 128, Akademiai Kiado ZRt., no. 3, pp. 1667–1688, viewed 7 May 2023, <https://link.springer.com/article/10.1007/s11192-022-04570-9>.
336. Vattai, ZA & Mályusz, L 2020, 'Negative weights in network time model', *Frontiers of Engineering Management 2020 9:2*, vol. 9, Springer, no. 2, pp. 268–280, viewed 6 May 2023, <https://link.springer.com/article/10.1007/s42524-020-0109-1>.
337. Vázquez-Serrano, JI, Peimbert-García, RE & Cárdenas-Barrón, LE 2021, 'Discrete-Event Simulation Modeling in Healthcare: A Comprehensive Review', *International Journal of Environmental Research and Public Health 2021, Vol. 18, Page 12262*, vol. 18, Multidisciplinary Digital Publishing Institute, no. 22, p. 12262, viewed 12 May 2023, <https://www.mdpi.com/1660-4601/18/22/12262/htm>.
338. Verganti, R, Vendraminelli, L & Iansiti, M 2020, 'Innovation and Design in the Age of Artificial Intelligence', *Journal of Product Innovation Management*, vol. 37, John Wiley & Sons, Ltd, no. 3, pp. 212–227, viewed 5 May 2023, <https://onlinelibrary.wiley.com/doi/full/10.1111/jpim.12523>.
339. Visser, M, van Eck, NJ & Waltman, L 2021, 'Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic', *Quantitative Science Studies*, vol. 2, MIT Press, no. 1, pp. 20–41, viewed 13 May 2023, <https://direct.mit.edu/qss/article/2/1/20/97574/Large-scale-comparison-of-bibliographic-data>.

340. Viti, M, Löwe, R, Sørup, HJD, Rasmussen, M, Arnbjerg-Nielsen, K & McKnight, US 2022, ‘Knowledge gaps and future research needs for assessing the non-market benefits of Nature-Based Solutions and Nature-Based Solution-like strategies’, *Science of The Total Environment*, vol. 841, Elsevier, p. 156636.
341. Van Vleck, TT, Farrell, D & Chan, L 2022, ‘Natural Language Processing in Nephrology’, *Advances in Chronic Kidney Disease*, vol. 29, W.B. Saunders, no. 5, pp. 465–471.
342. Vrchota, J, Řehoř, P, Maříková, M & Pech, M 2020, ‘Critical Success Factors of the Project Management in Relation to Industry 4.0 for Sustainability of Projects’, *Sustainability 2021, Vol. 13, Page 281*, vol. 13, Multidisciplinary Digital Publishing Institute, no. 1, p. 281, viewed 5 May 2023, <https://www.mdpi.com/2071-1050/13/1/281/htm>.
343. Vrontis, D, Christofi, M, Pereira, V, Tarba, S, Makrides, A & Trichina, E 2021, ‘Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review’, *https://doi.org/10.1080/09585192.2020.1871398*, vol. 33, Routledge, no. 6, pp. 1237–1266, viewed 5 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/09585192.2020.1871398>.
344. Wafa, R, Khan, MQ, Malik, F, Abdusalomov, AB, Cho, YI & Odarchenko, R 2022, ‘The Impact of Agile Methodology on Project Success, with a Moderating Role of Person’s Job Fit in the IT Industry of Pakistan’, *Applied Sciences 2022, Vol. 12, Page 10698*, vol. 12, Multidisciplinary Digital Publishing Institute, no. 21, p. 10698, viewed 5 May 2023, <https://www.mdpi.com/2076-3417/12/21/10698/htm>.
345. Wang, T 2020, ‘A dynamic model for evaluating the success of construction mega projects in China’, Hong Kong Polytechnic University, viewed 20 April 2023, <https://theses.lib.polyu.edu.hk/handle/200/10735>.
346. Wawak, S, Ljevo, Ž & Vukomanović, M 2020, ‘Understanding the Key Quality Factors in Construction Projects—A Systematic Literature Review’, *Sustainability 2020, Vol. 12, Page 10376*, vol. 12, Multidisciplinary Digital Publishing Institute, no. 24, p. 10376, viewed 4 May 2023, <https://www.mdpi.com/2071-1050/12/24/10376/htm>.
347. Wen, S, Tang, H, Ying, F & Wu, G 2023, ‘Exploring the Global Research Trends of Supply Chain Management of Construction Projects Based on a Bibliometric Analysis: Current Status and Future Prospects’, *Buildings 2023, Vol. 13, Page 373*, vol. 13, Multidisciplinary Digital Publishing Institute, no. 2, p. 373, viewed 7 May 2023, <https://www.mdpi.com/2075-5309/13/2/373/htm>.
348. Whyte, J n.d., ‘How Digital Information Transforms Project Delivery Models’.
349. Wilson, JM 2003, ‘Gantt charts: A centenary appreciation’, *European Journal of Operational Research*, vol. 149, North-Holland, no. 2, pp. 430–437.
350. Wu, C, Li, X, Jiang, R, Guo, Y, Wang, J & Yang, Z 2023, ‘Graph-based deep learning model for knowledge base completion in constraint management of construction projects’, *Computer-Aided Civil and Infrastructure Engineering*, vol. 38, John Wiley & Sons, Ltd, no. 6, pp. 702–719, viewed 12 May 2023, <https://onlinelibrary.wiley.com/doi/full/10.1111/mice.12904>.
351. Wu, H, Zuo, J, Zillante, G, Wang, J & Yuan, H 2019, ‘Construction and demolition waste research: a bibliometric analysis’, *https://doi.org/10.1080/00038628.2018.1564646*, vol. 62, Taylor & Francis, no. 4, pp. 354–365, viewed 7 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/00038628.2018.1564646>.
352. Xiao, B 2023, ‘Copyright law and non-fungible tokens: experience from China’, *International Journal of Law and Information Technology*, Oxford University Press (OUP).
353. Xu, S, Zhang, X, Feng, L & Yang, W 2020, ‘Disruption risks in supply chain management: a literature review based on bibliometric analysis’, *https://doi.org/10.1080/00207543.2020.1717011*, vol. 58, Taylor & Francis, no. 11, pp. 3508–3526, viewed 13 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/00207543.2020.1717011>.
354. Xu, Z, Ge, Z, Wang, X & Skare, M 2021a, ‘Bibliometric analysis of technology adoption literature published from 1997 to 2020’, *Technological Forecasting and Social Change*, vol. 170, North-Holland, p. 120896.

355. Xu, Z, Wang, Xindi, Wang, Xinxin & Skare, M 2021b, ‘A comprehensive bibliometric analysis of entrepreneurship and crisis literature published from 1984 to 2020’, *Journal of Business Research*, vol. 135, Elsevier, pp. 304–318.
356. Xue, R, Baron, C, Esteban, P & Zheng, L n.d., ‘Analysis and Comparison of Project Management Standards and Guides’.
357. Yeboah-Boateng, EO & Nwolley, S 2019, ‘Developing use cases for big data analytics: Data integration with social media metrics’, *Handbook on ICT in Developing Countries: Next Generation ICT Technologies*, River Publishers, pp. 171–199, viewed 7 May 2023, <https://www.taylorfrancis.com/chapters/edit/10.1201/9781003346043-8/developing-use-cases-big-data-analytics-data-integration-social-media-metrics-ezer-osei-yeboah-boateng-stephane-nwolley>.
358. Yew, ANJ, Schraagen, M, Otte, WM & van Diessen, E 2023, ‘Transforming epilepsy research: A systematic review on natural language processing applications’, *Epilepsia*, vol. 64, John Wiley & Sons, Ltd, no. 2, pp. 292–305, viewed 10 May 2023, <https://onlinelibrary.wiley.com/doi/full/10.1111/epi.17474>.
359. Zasa, FP, Patrucco, A & Pellizzoni, E 2021, ‘Managing the Hybrid Organization: How Can Agile and Traditional Project Management Coexist?’, *https://doi.org/10.1080/08956308.2021.1843331*, vol. 64, Routledge, no. 1, pp. 54–63, viewed 5 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/08956308.2021.1843331>.
360. Zayat, W & Senvar, O 2020, ‘Framework Study for Agile Software Development Via Scrum and Kanban’, *International Journal of Innovation and Technology Management*, vol. 17, World Scientific, no. 4.
361. Zhang, L, Mohandes, SR, Tong, J, Abadi, M, Banihashemi, S & Deng, B 2023, ‘Sustainable Project Governance: Scientometric Analysis and Emerging Trends’, *Sustainability 2023, Vol. 15, Page 2441*, vol. 15, Multidisciplinary Digital Publishing Institute, no. 3, p. 2441, viewed 4 May 2023, <https://www.mdpi.com/2071-1050/15/3/2441/htm>.
362. Zhang, Q, Wang, W & Chen, Y 2020, ‘Frontiers: In-Consumption Social Listening with Moment-to-Moment Unstructured Data: The Case of Movie Appreciation and Live Comments’, *https://doi.org/10.1287/mksc.2019.1215*, vol. 39, INFORMS, no. 2, pp. 285–295, viewed 7 May 2023, <https://pubsonline.informs.org/doi/abs/10.1287/mksc.2019.1215>.
363. Zhang, Y & Chen, Y 2020, ‘Research trends and areas of focus on the Chinese Loess Plateau: A bibliometric analysis during 1991–2018’, *CATENA*, vol. 194, Elsevier, p. 104798.
364. Zhao, M, Zhou, J, Wang, K & Pantelous, AA 2023, ‘Project scheduling problem with fuzzy activity durations: A novel operational law based solution framework’, *European Journal of Operational Research*, vol. 306, North-Holland, no. 2, pp. 519–534.
365. Zheng, X, Lu, J & Kiritsis, D 2021, ‘The emergence of cognitive digital twin: vision, challenges and opportunities’, *https://doi.org/10.1080/00207543.2021.2014591*, vol. 60, Taylor & Francis, no. 24, pp. 7610–7632, viewed 12 May 2023, <https://www.tandfonline.com/doi/abs/10.1080/00207543.2021.2014591>.
366. Zhou, C, Li, Q, Li, C, Yu, J, Liu, Y, Wang, G, Zhang, K, Ji, C, Yan, Q, He, L, Peng, H, Li, J, Wu, J, Liu, Z, Xie, P, Xiong, C, Pei, J, Yu, PS & Sun, L 2023, ‘A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT’, viewed 10 May 2023, <https://arxiv.org/abs/2302.09419v3>.
367. Zhu, J & Liu, W 2020, ‘A tale of two databases: the use of Web of Science and Scopus in academic papers’, *Scientometrics*, vol. 123, Springer Netherlands, no. 1, pp. 321–335, viewed 13 May 2023, <https://link.springer.com/article/10.1007/s11192-020-03387-8>.
368. Zulch, B 2014, ‘Leadership Communication in Project Management’, *Procedia - Social and Behavioral Sciences*, vol. 119, Elsevier, pp. 172–181.
369. Žužek, T, Gosar, Ž, Kušar, J & Berlec, T 2020, ‘Adopting Agile Project Management Practices in Non-Software SMEs: A Case Study of a Slovenian Medium-Sized Manufacturing Company’, *Sustainability 2020, Vol. 12, Page 9245*, vol. 12, Multidisciplinary Digital Publishing Institute, no. 21, p. 9245, viewed 5 May 2023, <https://www.mdpi.com/2071-1050/12/21/9245/htm>.