Risk analysis of implementing Machine Learning in construction projects

Aki Roy
Abstract

Machine Learning has significantly influenced development across domains by leveraging incoming and existing data. However, despite its advancements, criticism persists regarding its failure to adequately address real-world problems, with the construction domain being an example. Construction sector is crucial for global economic growth, yet it remains largely unexplored, lacking sufficient research and technological utilization of its extensive data. Despite increasing publications on adapting technological advancements, the primary focus is on urging industry innovation through digitization. Recently, adopting Machine Learning to address operational challenges has gained attention. While some studies have explored potential ML integration opportunities in construction, there is a gap in understanding the factors and barriers hindering its adoption across projects. This study investigates the factors restricting organizations from implementing ML in construction projects and their consequent operational impacts.

This study employs a comprehensive literature review of ML concepts and identifies gaps in construction data. Qualitative interviews have been conducted in a semi-structured manner with five industry professionals offering practical insights, preceding a thematic analysis of interview data. Themes are analyzed and discussed in relation to theoretical material to identify connections. Finally, a risk assessment based on identified risks is evaluated through a risk matrix. The results of this study discuss the challenges and potential benefits of implementing ML within the construction industry. The study further emphasizes the necessity of knowledge to understand project-specific datasets. With a primary focus on unstructured text and image data, the study uncovers challenges related to data inconsistency that affect data reliability. While recognizing ML’s potential to streamline construction operations, the study underscores challenges such as data security and digitalization. In summary, this study emphasizes the importance of data quality, security, and cultural transformation in harnessing ML’s capabilities to improve construction project management and operations.

Keywords: Construction, Machine Learning, Unstructured Data, Image Processing, Text Processing, Project Analysis, Data Management, Risk Identification.
Synopsis

Background
Artificial Intelligence and Machine Learning offer tailored applications across fields in this data era yet face challenges in real-world complexities, like in the construction domain. The construction industry significantly contributes to global economic growth, with a demand to digitize its business operations for more efficient project management. However, the available data often needs to be utilized more, leading to low efficiency and backward management. This industry is characterized by high levels of uncertainty and complexity, making risk assessment and control difficult. It has been suggested that Machine Learning be implemented in construction projects to decrease risk and increase control with significant improvements in project management, particularly in cost estimation and prediction. Construction projects can become "smart" and optimize resources by incorporating technologies such as ML. A thorough analysis of the current trends in ML in construction management is crucial for recommending new research directions.

Problem
There is a lack of comprehensive literature review on the use of ML in construction project management, as limited empirical studies have examined the problems that restrict its implementation. While previous studies have emphasized the benefits of using ML in construction projects, there needs to be more real industrial cases that have either successfully implemented or failed in the implementation process. This knowledge gap leads to a limited understanding of the risks associated with implementing ML in construction projects, which can result in higher costs, project delays, and reduced sustainability. This study addresses the problem of limited empirical studies on the challenges and hindrances associated with using ML in construction projects.

Research Question
The paper adopts a qualitative approach to research ML in construction projects. It aims to answer research questions that seek to contribute to the academic and practical concepts of ML by providing insights into the current state and future
possibilities of ML implementation in construction projects and identifying the challenges and opportunities associated with its use. To answer these questions, three research questions have been formulated: (RQ1) What are the main risks and challenges of implementing ML in construction projects, and how do they impact construction operations?, (RQ2) How can ML contribute to value creation in operations compared to a project-level analysis?, (RQ3) How is ML beneficial for the future of projects in the construction sector?.

Method
This thesis uses a case study approach to comprehensively understand the problem and its context. The data collection method is in the form of interviews, documents, and data analysis methods in the form of thematic analysis to define and evaluate risk analysis based on the perceived information from the stakeholders within the organization. The individuals selected for the interviews were stakeholders from a private company with prior knowledge and technical expertise in ML and data management to ensure the most authentic information regarding ML implementation. The study aimed to assist stakeholders, project managers, and affected individuals in implementing ML and taking the necessary precautions for a successful project. Risk analysis findings were used to determine the best way to utilize ML in an organization.

Result
Five themes have been identified from respondents’ answers: (1) data handling, which helps to understand how to interpret the existing construction data. On the contrary, this data has errors leading to deteriorated data reliability; (2) technological advancement, focuses on using ML to serve better data usage for produced construction data, but due to the uniqueness of each project, ML is complicated to implement in every project; (3) ML implementation necessities, help identify what is required for implementation, whether organizational structure change or understanding data algorithms. However, it only covers some implementation necessities as every project is individual and requires change by need; (4) Benefits with ML implementation, explains how an ML implementation reduce workload, enhance efficiency, analyze root cause analysis, examine safety data, support standardization, and facilitate decision-making; and (5) Challenges with ML implementation, prepare to be aware of the implementation issues that can be mitigated along the way; however, these identified challenges are not the only ones and cannot be solely relied upon.

Discussion
The construction industry faces challenges in adopting ML due to project-specific data variation. ML can significantly improve project management, but risks include data quality issues, resistance to new technology, and limited data sharing. Partial ML implementation due to data constraints requires comparing
datasets across projects for comprehensive analysis, emphasizing the need for a problem-centric approach linking ML research to real-world application. ML enhances predictive maintenance, optimizes resource allocation, and refines processes through data analysis to create operational value. It also improves image and text processing, project cost estimation, and overall decision-making. The future of construction projects relies on efficiently harnessing ML to analyze vast amounts of project data, enhance cross-project analysis, identify hazards, and foster collaboration among stakeholders and technology experts.
Acknowledgement

I would like to extend my deepest gratitude and appreciation to all those who have contributed to successfully completing my master’s thesis. This work would not have been possible without their unwavering support, guidance, and encouragement.

First and foremost, I am immensely grateful to my supervisor, Oskar Juhlin, for his invaluable guidance and mentorship throughout this research journey. Oskar’s expertise and unwavering support have played a pivotal role in shaping the direction of this thesis and expanding my understanding of the subject matter. I would also like to sincerely thank my industry supervisors at PlanB, Carl Henrikson and Dan Engström, for their insightful feedback and valuable suggestions. Their expertise and scholarly input have significantly enriched the quality of this thesis. Additionally, I would like to acknowledge the participants and individuals who graciously devoted their time and expertise to participate in the interviews for data collection. Their contributions have been invaluable; this research would not have been possible without their involvement.

I want to express my heartfelt appreciation to my parents for their unwavering love, encouragement, and belief in my abilities. Their patience, understanding, and sacrifices have been the bedrock of my academic pursuits, and I am profoundly grateful for their constant support. Lastly, I would like to thank my husband, who has supported me throughout this journey. His encouragement, insights into the thesis topic, and persistent presence have been instrumental in my success.
# Contents

List of Figures .......................................................... ii  
List of Tables ............................................................ iii  
List of Abbreviations ...................................................... iv  

1 Introduction ................................................................ 1  
1.1 Background ............................................................ 1  
1.2 Problem ................................................................. 3  
1.3 Research question ...................................................... 4  
1.4 Delimitation ............................................................. 4  

2 Literature background ................................................. 6  
2.1 Machine Learning in construction projects ...................... 6  
2.1.1 Research contribution ............................................ 8  
2.2 Problem in Machine Learning ..................................... 9  
2.2.1 Challenges with Machine Learning in construction .... 10  
2.3 Construction data ..................................................... 10  
2.4 Construction project lifecycle ..................................... 12  
2.5 Risk management in construction projects .................... 13  

3 Methodology ................................................................ 15  
3.1 Research strategy ....................................................... 15  
3.1.1 Case study .......................................................... 15  
3.2 Data collection method ................................................. 16  
3.2.1 Semi-structured interview ..................................... 16  
3.2.2 Document analysis ............................................... 16  
3.3 Data analysis method .................................................. 17  
3.3.1 Thematic analysis ................................................ 17  
3.4 Selection of interviewees .............................................. 17  
3.4.1 Respondent profile ............................................... 18  
3.5 Interview guide ........................................................ 18  
3.6 Ethical consideration .................................................. 19  
3.7 Alternate strategy and methods .................................... 20
4 Results
4.1 Identified themes ........................................... 22
  4.1.1 Data handling ........................................... 22
  4.1.2 Technological advancement ........................ 24
  4.1.3 ML implementation necessities .................... 25
  4.1.4 Benefits with ML implementation .................. 27
  4.1.5 Challenges with ML implementation ............... 28

5 Risk identification ........................................... 31
  5.1 Risk identification ........................................ 31
  5.1.1 Risk evaluation ........................................ 32
  5.1.2 Risk matrix ............................................ 32

6 Discussion and conclusion .................................. 33
  6.1 Discussion ............................................... 33
  6.2 Conclusion .............................................. 35
  6.3 Limitations .............................................. 36
  6.4 Ethical and societal consequences .................... 36

7 Future research ............................................. 38

8 Bibliography ............................................... 39

Appendices .................................................. 42

9 Consent form ............................................. 43

10 Reflection Document ..................................... 45
# List of Figures

1.1 Annual ML publication in construction management (Wang, 2021) 3  
2.1 Project fragmentation (Alashwal and Hamzah, 2014) . . . . . . . 12  
2.2 Construction project fragmentation (Sawhney et al., 2020) . . . . 13  
2.3 Risk structure in construction (Zavadskas et al., 2010) . . . . . . 14  
4.1 Core theme 1 “Data handling” . . . . . . . . . . . . . . . . . . . . . 22  
4.2 Core theme 2 “Technological advancement” . . . . . . . . . . . . 24  
4.3 Core theme 3 “ML implementation necessities” . . . . . . . . . . . 25  
4.4 Core theme 4 “Benefits with ML implementation” . . . . . . . . . 27  
4.5 Core theme 5 “Challenges with ML implementation” . . . . . . . 28  
5.1 Risk matrix and the 4Ts of hazard management . . . . . . . . . . 32
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Respondent profile</td>
<td>18</td>
</tr>
<tr>
<td>3.2</td>
<td>Interview guide</td>
<td>19</td>
</tr>
<tr>
<td>5.1</td>
<td>Risk identification</td>
<td>31</td>
</tr>
<tr>
<td>5.2</td>
<td>Risk evaluation</td>
<td>32</td>
</tr>
</tbody>
</table>
List of Abbreviations

AI - Artificial Intelligence
ML - Machine Learning
GDP - Gross Domestic Product
EU - European Union
BIM - Building Information Modelling
GDPR - General Data Protection Regulation
SMEs - Small and medium-sized enterprises
CAD - Computer-aided design
KPI - Key Risk Factor
Chapter 1

Introduction

The first chapter introduces the context of the research work. It elaborates on the problem formulation, research question(s), and limitations of the study.

1.1 Background

We currently live in the data era, where everything is linked to a data source, and our daily experiences are documented digitally. Analyzing these data holds the potential to create diverse, intelligent applications tailored to specific fields. Artificial intelligence (AI), especially machine learning (ML), has experienced rapid development in recent years, primarily in the area of data analysis and computation, enabling applications to operate intelligently (Sarker, 2021). ML systems are quickly expanding in size and capabilities and are increasingly utilized in critical scenarios. However, ML systems must be resilient to unexpected events and uncommon challenges to function effectively in high-stakes real-world environments (Hendrycks et al., 2021). Various ML techniques are employed across various real-world applications, including cybersecurity systems, smart cities, healthcare, e-commerce, agriculture, and numerous other domains (Sarker, 2021). Unfortunately, current ML systems often struggle with the intricacies of real-world complexities and unforeseeable factors (Hendrycks et al., 2021), and one such industry is the construction industry.

The construction industry is a large sector that plays a prominent role in economic growth with an increasing demand to digitize its business by effectively transforming its existing operations. This industry is data intensive, where the available data is only partially utilized. This leads to backward management and low efficiency in the construction sector as the project schedule, costs, and quality need to be more effectively controlled (You and Wu, 2019). A survey from the Mckinsey Global Institute in 2017 states that the global construction industry represents around 13% of the world’s Gross Domestic Product (GDP) (Pan and Zhang, 2021). Consecutively, construction generates about 9% of gross domestic products in the European Union (EU) and provides 18 million
direct jobs, which are generally project-based. These project-based operations are characterized by ad-hoc, with varying goals and methods. Construction projects are characterized by high levels of uncertainty and complexity, which means that conditions can shift unpredictably at any time. As a result, practitioners need to reconsider their methods for assessing and controlling risks within their organizations and projects (Basaif et al., 2020). Moreover, construction projects involve activities and processes requiring input from various disciplines, particularly when adopting Building Information Modelling (BIM) technologies. However, this has led to several challenges, including data inconsistency, version control, and ambiguity in the data ownership. Large volumes of data are generated during these projects, which must be carefully managed and updated throughout the project lifecycle. Although collaboration tools and data servers like ProjectWise and RevitServer have been developed to facilitate this process, developing companies often control their data governance approach for commercial and competitive reasons (Alreshidi et al., 2014).

The term "data-driven transformation" pertains to organizational endeavors that optimize data utilization, often by incorporating new technologies, to achieve various objectives such as revenue forecasting or improving operational efficiency. The advantages of utilizing data to drive decision-making and transformation are widely recognized across industries and can improve business performance and operational efficiency (Wang, 2021). The construction sector relies heavily on data, but the small datasets available in this sector need to be fully utilized, otherwise, it might result in poor management and inefficient processes. Furthermore, construction projects entail complex activities and processes that require collaboration across different disciplines, especially when implementing Building Information Modelling (BIM) technologies (Tambe et al., 2019). Most project-based operations act individually and require collecting more data points for the current workforce. By relying on theory and prior research, organizations can uncover causal relationships and predictors of outcomes, even if the data set needs to be larger to build an algorithm. At the same time, even if data sets need to be bigger to learn more, small data sets can still be effective in identifying relationships, but we may not be able to build an algorithm (ibid). Depending on the use case, this can involve traditional analytics, like descriptive statistical analysis, or advanced algorithms, such as AI/ML. Figure 1.1 illustrates the rapid increase in the number of relevant publications over the past 21 years. The upward trend in the number of relevant publications from 2016 to 2020 shows the importance of using ML in the construction industry (Wang, 2021).
ML plays a vital role in the benefit of the construction industry, as it has the potential to enhance project productivity and quality and even has an academic advantage which is necessary for researchers and professionals. Despite appearing to be a far-off concept that may take decades to materialize, the future of this technology is much nearer. In recent years, ML has gained attraction in the construction sector, with significant improvements in project management in different areas, such as cost estimation, prediction, etc., (Van and Toan, 2021). By incorporating technologies like ML, construction projects can become "smart" which is valuable, particularly when dealing with substantial volumes of data, as they save time and optimize resources (Xu et al., 2021). Future studies in ML for construction management will integrate and categorize existing literature reviews to enhance comprehension and facilitate more in-depth and efficient research. Hence, conducting a thorough analysis of current trends in ML in construction management is crucial to recommend new research directions (Van and Toan, 2021).

1.2 Problem

The research topic in construction project management concerning ML is highly diversified, but a comprehensive literature review on the subject is lacking (Van and Toan, 2021). The problem that this thesis addresses is that prior studies mainly highlight the benefits of using ML in construction projects, however, only limited empirical studies examine the problems that restrict the use of ML in such projects. This contributes to limited examples of real industrial cases that have either successfully implemented ML or failed in the implementation process, leading to restricted knowledge about its risks. The industry’s limited adoption of analyzing the risks of implementing ML in construction projects may lead to various adverse outcomes, such as valuable data could be buried.
in large volumes of records due to difficulties in extracting files on time. Moreover, many information-intensive historical data could be left untouched after projects are completed without further examination to improve future project procedures (Soibelman et al., 2008). Everyday construction projects are creating high amount of data but are suffering with limited analyzing capability with advancements like ML. To address the research problem, this study needs to analyze the factors prohibiting organizations from implementing ML in construction projects and the impact on construction operations. This result will benefit project managers, investors, and other relevant stakeholders to understand and add value to current and future projects and provide more effective and sustainable technological solutions.

1.3 Research question

The research questions aim to address the issue at hand and make both academic and practical contributions to the field of ML, particularly in its real-world applications. The research problem pertains to an aspect of the construction industry that has received little attention for a long time. This research area needs technological enhancement through solutions like ML to align with today’s advancements. The formulated research questions will help understand the technological gap and facilitate mapping down the requirements to implement ML in future projects. This paper undertakes a qualitative approach to explore the integration of ML in construction projects with the following objectives:

**RQ1.** What are the main risks and challenges of implementing ML in construction projects, and how do they impact construction operations?

**RQ2.** How can ML contribute to value creation in operations compared to a project-level analysis?

**RQ3.** How is ML beneficial for the future of projects in the construction sector?

1.4 Delimitation

This study is constrained by a few limitations that affect its ability to obtain relevant interviewees for qualitative research. Firstly, the study is limited to candidates with general knowledge of ML, narrowing the scope of the target group. Most interviewees do not have practical knowledge of ML implementation in projects, which challenges the reliability and validity of the qualitative data. To understand the data, the interviewees may have retrieved it from open AI platforms, which may not represent real-life research variables. Additionally, the study faces difficulties in understanding the patterns of construction datasets since each project has its unique data based on project requirements. Another limitation is using relevant keywords to retrieve publications on ML in construction projects, which may only encompass some possible terms. This
could affect the research outcomes and might not entirely represent the entire literature accessible in this domain.
Chapter 2

Literature background

This chapter provides a frame of reference for the current study. This includes presenting relevant works and references in the field and discussing related terms. Furthermore, the chapter delves into using ML in construction projects and its problems, highlighting its unique characteristics.

2.1 Machine Learning in construction projects

AI is a fundamental technology in the context of the Industrial Revolution 4.0, and within this field, machine learning has emerged as a specialized subset. This interdisciplinary domain can learn from data and is designed to handle situations where there is uncertainty or incomplete information, and it can use past experiences to make decisions in similar scenarios, as stated by Nguyen (2021). ML can be categorized into two approaches: supervised and unsupervised learning. The supervised approach is typically used for regression or classification tasks, whereas unsupervised learning is employed for clustering or dimension-reduction tasks. Semi-supervised learning is another form of ML that combines both supervised and unsupervised learning methods. This approach only requires target values for a subset of the data used, allowing the analysis of incomplete datasets and sometimes achieving better results than classical supervised learning methods. However, assumptions about distribution densities must be made in advance for semi-supervised learning algorithms. If these assumptions are unfavorable, the results can be significantly worse than with a supervised learning procedure (Schmid et al., 2021). ML allows information about a system to be learned from observed data without relying on predetermined mechanistic relationships. This is achieved by generating multiple instances with varying initial conditions and training them in parallel. The performance of the ML model is then evaluated on a set of several thousand instances of the same model rather than just a single instance. The optimal design is determined by comparing the average performance between runs of various ML models, even if individual instances cannot be compared. Once the
optimal design is selected, the best-performing instance of that design is used as the final model (Shaikhina et al., 2015).

ML has the remarkable ability to anticipate potential hazards and assist people in identifying and avoiding risks. This feature has led to numerous studies examining the application of ML in predicting and evaluating the dangers associated with construction projects (Van and Quoc, 2021). Project management is known for its fluid and flexible nature, and construction project analytics presents an opportunity to apply ML algorithms, where ML belongs to the category of cognitive analytics, which focuses on predicting problems that can occur during the project. Researchers have been investigating ways to use large amounts of project data in construction projects. ML techniques have concentrated explicitly on looking back at past reports and grasping the fundamental relationships to enable informed decision-making, like streamlining the design of construction projects, for instance (Uddin et al., 2022).

ML has the remarkable ability to anticipate potential hazards and assist people in identifying and avoiding risks. This feature has led to numerous studies examining the application of ML in predicting and evaluating the dangers associated with construction projects (Van and Quoc, 2021). Project management is known for its fluid and flexible nature, and construction project analytics presents an opportunity to apply ML algorithms, where ML belongs to the category of cognitive analytics, which focuses on predicting problems that can occur during the project. Researchers have been investigating ways to use large amounts of project data in construction projects. ML techniques have concentrated explicitly on looking back at past reports and grasping the fundamental relationships to enable informed decision-making, like streamlining the design of construction projects, for instance (Uddin et al., 2022).

Uddin et al. (2022), conducted a study evaluating existing ML methods in construction project delivery. The study introduced an ML-based data-driven framework to address project analysis issues, offering significant contributions to project practitioners, stakeholders, and academics in resolving various project-related challenges. Another relevant study, "A risk prediction paradigm based on a cross-analytical machine learning model" by Banerjee et al. (2021), underscores the persistent challenge of identifying and timely avoiding inherent risks, particularly in larger construction projects referred to as "megaprojects". Numerous studies have looked into risk management within construction projects, incorporating ML algorithms to cluster risk factors based on their interdependencies and cumulative impact on project performance. Recognizing those risks vary across projects due to their unique nature, distribution, stakeholders, and investment, assessing risks tailored to each project. Political, design and economic risks are predominant, and unforeseen losses significantly impact construction projects, leading to delays and financial losses. Apart from the mentioned risk factors, environmental, execution, and social risks influence project success, as illustrated in Figure 2.3. Examining these risks reveals a lack of a comprehensive and uniform classification approach that considers interdependencies among risk factors and their collective impact on overall project success (Banerjee et al., 2021).
Banerjee et al. (2021), underline several benefits in their study, including identifying Key Risk Factors (KPIs) and providing valuable insights for project managers and stakeholders to anticipate and tackle potential challenges. It also offers a comprehensive definition of risk in the context of construction projects, emphasizing the significance of understanding how risks impact cost-benefit analysis, local perceptions, construction costs, and other financial variables for well-informed decision-making. On the other hand, by pinpointing risk factors in the study and their potential consequences, the research advocates for specific management actions to prevent prolonged delays, financial losses, and conflicts among stakeholders, ultimately aiming to prevent project failure. However, the study states that qualitative or quantitative approaches may not provide a generalized risk management framework for identifying risk factors. While quantitative methods or analytical approaches incorporating expert responses to risk factors can help identify significant risks, they may fall short in addressing sub-risks that collectively impact project success. Banerjee et al. (2021), research revealed that insufficient preliminary surveys appear to be a significant risk factor influencing the scope of construction projects. To achieve target performance, it is crucial to have better reachability or connectivity and procure high-quality materials. The study emphasizes that current methods lack the ability to contribute to a project-specific risk management framework and are, therefore, not universally applicable to all projects. Banerjee et al. (2021), surpass these limitations by clarifying that conducting risk assessments at the beginning of a project can identify potential losses and persistent issues in construction projects. Therefore, evaluating various risks, their interrelationships, and underlying driving forces and implementing corresponding mitigation ensures their significance (Banerjee et al., 2021).

2.1.1 Research contribution

Despite increasing publications of literature on technological advancements in construction, most research only emphasizes the need for the construction industry to seek innovative ways to improve productivity, with digitization being a significant driver of such efficiency on construction sites, where data-driven decision-making has proven to be effective in promoting sustainable processes (Erikshammar and Steln, 2022). However, there is a gap in research addressing whether construction projects should embrace solutions like ML and how it shall prepare and decide on such implementations. While ML holds great promise for addressing operational issues such as safety, productivity, and quality in the construction sector, it fails to argue if a project is viable for such an implementation and the associated risks in real-life industrial cases (KarimiAzari et al., 2011; Van and Quoc, 2021).

Although some studies have explored the potential opportunities for integrating ML in various aspects of construction, studies have yet to explore the factors and barriers preventing organizations from adopting this technological advancement in the domain. Furthermore, the construction industry lacks substantial knowledge of risk assessment, which is particularly crucial given the industry’s
inherent uncertainties, hence, it should be a crucial aspect of decision-making for construction companies (KarimiAzari et al., 2011). Banerjee et al. (2021) note that despite the rapid growth in the global construction sector and the probability of risks, more studies need to focus on risk management in construction projects. Therefore, grouping various risk factors according to their connection can aid in categorizing primary risks and sub-risk factors. This approach, in turn, will facilitate more effective distribution and management of risks, ultimately ensuring the success of a project (Banerjee et al., 2021). On the other hand, as each project is unique and is specific to its results, investigating the perspectives of industry professionals familiar with ML is necessary to gain insights into the advantages and disadvantages of this technology and how its future prospects and implementation may impact areas of construction projects that could benefit from digitization and contribute to future research in construction projects.

2.2 Problem in Machine Learning

Wagstaff (2012) mentions a few notable challenges with ML implementation in "Machine Learning that Matters". The paper summarizes that despite its advantages, ML needs to be more focused on solving real-world practical problems important for science and society. Although ML highlights identifying problems, collecting data, choosing learning methods, and involving domain experts, the importance of these activities in achieving meaningful results, developing an algorithm, or choosing metrics viable for experiments are typically considered "publishable" in the ML community. This emphasis that algorithm development or data analysis often overlooks the practical application and impact on the outside world (Wagstaff, 2012). Wagstaff (2012) also mentioned that the data generated for ML is subjective as the researchers can choose the data they study, the criteria they use to measure how well their models work, and how they share the results with the specific domain they work for. Using familiar datasets offers potential benefits, such as facilitating direct empirical comparisons with other methods and making it easier to interpret results, assuming the dataset’s properties are well-known. However, practical challenges arise as direct comparisons often need to be standardized for reproducibility (Wagstaff, 2012).

Other challenges include using ML models in concepts and language that are difficult to understand and act as communication barriers (Wagstaff, 2012). The three common issues in ML language include: (1) suggestive definitions, (2) overloaded terminology, and (3) suitcase words. New technical words are often created with casual meanings, leading to confusion as subsequent papers either embrace or replace these terms without proper explanation (Lipton and Steinhardt, 2018). To overcome this, the suggestion is to express ML concepts using more general terms familiar to broader ML models, facilitating better understanding (Wagstaff, 2012). Another obstacle highlighted is the increased risk associated with ML system responsibility and adjustments. Despite an
ML system having an error rate comparable to that of a human performing the same task, depending on the machine may evoke a sense of increased risk due to the emergence of new concerns. These concerns include who shares the responsibility for that error, given the growing size of the implementation. Thus, we need to address them before including ML in real systems. Other challenges also include the complexity and immaturity of the ML field, which has yet to reach a point where researchers from other domains can easily apply ML to their problems. The core reason is the lack of knowledge, which leads to the incapability to explain the problem, decide which features are essential, or figure out the correct settings for the ML process. In order to tackle this problem, efforts are needed to simplify, mature, and robustify ML algorithms and tools, making them more accessible to a broader range of users (Wagstaff, 2012).

2.2.1 Challenges with Machine Learning in construction

A study by Regona et al. (2022) signifies the generation of stored data by embracing AI within the construction industry. Establishing connections among all essential elements for integrating AI in construction will be necessary for a seamless shift from traditional approaches. Moreover, key drivers of AI applications in construction include (1) producing outcomes easily comprehensible to all stakeholders and (2) enhancing consistency and reliability, given AI’s high resistance to errors. On the other side, the pivotal challenges include (1) the nature of AI to continuously use algorithms to train its pattern, (2) potential data issues due to the fragmented nature of the construction industry, (3) incompatibility between existing construction processes and (4) the continuous need for investment to maintain up-to-date and accurate data. Additionally, other hindrances to the development of AI algorithms include (1) ensuring the security and reliability of large datasets, (2) the possibility of non-actionable tasks, and (3) the lack of standardization in construction projects, making AI implementation challenging. To overcome these challenges, construction companies must adapt their corporate structures to integrate AI. Regona et al. (2022) also mention that companies should identify areas where AI algorithms can have the most immediate impact. Without a clear business case, AI adoption produces risks with high impact through technologies that assess risk and prevent cost overruns using ML algorithms (Regona et al., 2022).

2.3 Construction data

Managing construction projects is a demanding and intricate task, as each project is unique and tailored to a specific client. Integrating data-driven solutions into construction management could significantly improve current practices, but implementing advanced technologies like deep learning faces various challenges. The most notable hurdle is the limited availability and accessibility of data in the construction industry. Many data recordings still need to be paper-based, and database systems are infrequently used, resulting in a short-
age of consolidated and organized data sources. This makes it difficult to create large datasets necessary for effective deep learning algorithms, which require extensive datasets to yield favorable results. Small datasets are problematic, and several techniques have been proposed to address this, including improving the training datasets, generating synthetic data from simulations, using pre-trained models from other datasets, and using generative models to produce new data. However, these methods are primarily effective for transformation invariant data like images and may not be suitable for transformation variant data like financial data (Delgado and Oyedele, 2021). Despite limited opportunities to extract value from existing small datasets, incorporating ML into construction projects poses various challenges, primarily obtaining labeled data, especially in complex construction projects (Xu et al., 2021). Due to limited datasets, unsupervised ML models can greatly enhance data analytics efficiency, particularly in scenarios with a shortage of training data. These models can enhance diverse datasets by retaining non-linear similarities in the input data and recreating them in the generated data. The construction industry commonly employs decentralized, unstructured, and unlabelled data sources that need more traditional data management and exchange methods to integrate distinct data types (Delgado and Oyedele, 2021).

Historically, construction data has focused on text-based, web-based, and image-based analysis using transactional databases, where each item is described in a data table or spreadsheet by a list of values associated with specific features. However, there are various other data sources in construction beyond those in spreadsheets. These encompass semi-structured or unstructured text documents like contracts, specifications, requests for information, and change orders, as well as unstructured multimedia data such as binary pictures, 2D or 3D drawings, and audio or video files. In construction projects, a significant proportion of project information is communicated using text documents, such as contracts, field reports, and requests for information, among others. However, current technologies for project document management do not offer direct support for integrating these various sources. This presents critical issues due to differences in vocabulary, large number of documents, and project model objects. While information systems used in construction have search engines based on term matching, using these tools may not always be effective, particularly in cases where multiple words share both the same and multiple meanings or relevant documents lack user-defined search terms (Soibelman et al., 2008).

Secondly, crucial details are embedded in the original data formats when working with construction image data. These include graphical representations of plans and textual information, which may not be preserved if the data is consolidated into a single data table or spreadsheet. Hence, it is necessary to categorize the numerous images meaningfully. Each image serves multiple functions, and even pictures taken for safety purposes may contain other valuable information that could be useful later, such as details about materials or structural elements. Besides image characteristics like materials, time data can be automatically extracted from digital cameras, such as timestamps when the images are captured at construction sites (Soibelman et al., 2008).
2.4 Construction project lifecycle

The construction industry is a multifaceted and constantly evolving sector that involves various stakeholders across different stages of planning, designing, constructing, and maintaining buildings. Fragmentation is a common issue within the industry, primarily stemming from separating the design and construction phases and the construction process itself. This has resulted in several challenges, including poor coordination between professionals and isolation between design and construction. The diagram in Figure 2.1 illustrates two levels of fragmentation: industry level, also known as firm fragmentation, and project level fragmentation. Industry-level fragmentation arises from segregating firms in the construction industry, resulting in numerous small and medium-sized enterprises (SMEs) and a decreasing average firm size. The majority of skills and knowledge lie with SMEs, but they face obstacles such as insufficient funding for research and development, as well as dynamic and temporary relationships with construction partners and different disciplines. This leads to complex contracts, low trust, and adversarial relationships between stakeholders. On the other hand, project-level fragmentation involves a lack of coordination, collaboration, and integration between functional disciplines and contractual partners during construction processes, resulting in inefficient work processes and poor communication. The separation between construction stages exacerbates project-level fragmentation. Collaboration and integration of professionals can mitigate project-level fragmentation and improve communication and knowledge sharing (Alashwal and Hamzah, 2014).

Figure 2.1: Project fragmentation (Alashwal and Hamzah, 2014)

According to Sawhney et al. (2020), the productivity index indicates that industry fragmentation is a long-running issue with structural challenges. Figure 2.2 shows the disjointed life cycles of typical construction projects leading to different fragmentations. Different types of fragmentation can complicate
the delivery process of construction projects. Vertical fragmentation happens when there are separate phases in a project, which requires multiple hand-offs of knowledge and can create risks. The complexity of the building design causes horizontal fragmentation and requires specialist knowledge from different disciplines to ensure safety. When the fragmentation fails to communicate or consider each other’s needs, problems can arise. Longitudinal fragmentation occurs when the knowledge gained from previous projects is not shared within the organization, hindering organizational learning and improvement. While knowledge management systems are available, they are rarely used in construction due to time constraints and the perception that each project’s problems are unique, and investments are often lost between projects due to fragmentation (Jones et al., 2022).

Figure 2.2: Construction project fragmentation (Sawhney et al., 2020)

2.5 Risk management in construction projects

Identifying and tracing the root causes of risks through a project to their consequences is the most challenging part of project management. In construction project management, risk management is a thorough and systematic approach that involves identifying, analyzing, and responding to risks to achieve project objectives. To succeed in construction projects, completing the project within a specific time, budget, and technical requirement is critical, and risk management is essential for achieving these objectives. Large construction projects are exposed to uncertainties due to various factors such as planning, design, construction complexity, involvement of multiple interest groups, and environmental factors (Banaitiene and Banaitis, 2012).

As construction projects can be unpredictable, risk management is vital for
achieving project objectives regarding time, cost, quality, safety, and environmental sustainability. Risk management is an iterative process that should be implemented systematically throughout a project’s lifecycle, from planning to completion, to help determine future consequences. A typical risk management process includes the following steps: risk identification, risk assessment, risk mitigation, and risk monitoring. Starting risk management in the early stages of a project is crucial, as it can influence major decisions such as alignment and construction methods (Banaitiene and Banaitis, 2012).

Construction projects are characterized by their uniqueness and the fact that they are typically built once, making the life cycle of these projects complete with various risks. These risks come from various sources, including temporary project teams from different companies, challenging conditions at construction sites, etc. Additionally, construction objects’ growing size and complexity further amplify these risks. The concept of “object risk” can be defined as an uncertain event or condition that, if it materializes, can positively or negatively impact one or more project objectives, such as time, cost, and quality. Evaluating these risks involves determining the level of risk associated with each objective and analyzing the potential risks using various methods and technologies (Zavadskas et al., 2010).

Project risks can be categorized into three main groups: (1) external risks, which are beyond the control of the project management team; (2) internal risks, which can be further subdivided based on the party responsible for the risk event, such as stakeholders, contractors, etc.; and (3) project risks, which directly impact the project’s completion (Zavadskas et al., 2010). In summary, the diagram below provides an overview of the various risks within the context of construction projects:

![Figure 2.3: Risk structure in construction (Zavadskas et al., 2010)](image-url)
Chapter 3

Methodology

This chapter explains the steps and methodologies utilized in the current study. The research process, data collection method, and ethical considerations are elaborately described.

3.1 Research strategy

3.1.1 Case study

Qualitative research provides a detailed description and understanding of a particular phenomenon. The approach taken in qualitative research is generally flexible and determined by the specific aims and nature of the study. A popular method within qualitative research is the case study approach. Case studies are useful when delving deeply into a specific topic to gain a greater understanding that may not be achievable through other methods (Njie and Asimiran, 2014). One of the benefits of the case study approach is its adaptability to a range of research methods, thus, it can be used for qualitative research. By focusing on a single case, researchers can uncover insights that may have wider implications and examine multiple factors, events, and relationships that occur in real-life (Denscombe, 2010).

This research strategy is well suited for this study because it enables a comprehensive investigation into the intricate complexities of real-life situations within the construction sector. Employing a case study approach will facilitate qualitative data collection through interviews to identify sub-risks, offering detailed and in-depth analysis. The case is to understand the current utilization of data in construction projects and explore ways to enhance its usability and scalability. Given the vast amount of data involved in the construction domain, there is a necessity for standardized digital methods to analyze and repurpose the data for future projects. The case study will investigate the feasibility of integrating ML into construction projects and examine its practical application within the construction context. The case study will consecutively help identify
ML implementation risks from the data, as it allows for incorporating multiple data sources from various construction projects involving different clients. Given the diverse ongoing projects in the construction sector, each with its unique variables, a case study provides the means to explore the relationships and complexities inherent in these real-world contexts. A survey may not adequately capture these complex aspects of ML implementation, which typically presents more generalized information with limited response options. Moreover, interviewees’ participation in different construction projects ensures varied examples and comparisons during interviews, although the questions assessing the risks will remain consistent across projects.

3.2 Data collection method

3.2.1 Semi-structured interview

The choice of data collection method depends on the research problem and the organization and analysis strategy of the research data. To ensure high research quality, it is essential to understand different methods and use the most suitable one to analyze the research question. Börjesson (2004) suggests that interviews are advantageous for collecting detailed and accurate data because of the relaxed structure of conducting interviews, with room for follow-up questions based on the research requirements.

This thesis study uses a qualitative method by collecting detailed data around the research questions using semi-structured interviews to understand the context of implementing ML and its associated risks. This method allowed an in-depth understanding of such implementation’s existing challenges and status in real projects. The research topic requires elaborative knowledge of how the technical aspects of ML and project data can be standardized, which is possible through a longer communication like an interview. Semi-structured interviews are considered a best practice for gathering detailed information, allowing participants to share in-depth information and researchers to gain valuable insights from the interviews (Denscombe, 2010). This method allowed altering the interview questions based on information and sequence directed by the participant. It even allowed the researcher to ask follow-up questions based on the participant’s previous answers, creating flexibility in data collection and the development of the interview.

3.2.2 Document analysis

Document analysis is a source to get research material due to the great accessibility through a library or the internet. For research studies, documentary sources should be evaluated to the criteria of (1) authenticity, (2) representativeness, (3) meaning, and (4) credibility (Denscombe, 2010). Document analysis is a method of researching similar topics that rely on existing literature as a source of information rather than raw data. This involves sifting through and compiling
relevant research material from previous studies. This approach is well-suited to
the current study, which requires a thorough risk analysis to evaluate the feasi-
bility of implementing ML in various projects. By examining the requirements
for ML implementation, the context of the study can be established.

Document analysis in this thesis study has considered materials such as
books, journals presented at conferences, official statistics, and company records.
These materials have been accessed mostly through Google Scholar by consid-
ering additional reliability by following the four criteria mentioned. The four
criteria analyzed the genuineness of articles, the typicality of the document,
clarity, and credibility of the document source (Denscombe, 2010).

3.3 Data analysis method

3.3.1 Thematic analysis

The compilation of data is carried out using *thematic analysis*, a qualitative
analysis method used to identify and analyze themes and patterns of collected
data. Thematic analysis helps explain similarities and differences between ap-
proaches with common features (Braun and Clarke, 2006). The thematic anal-
ysis helps to identify recurring patterns from the conducted interviews during
the evaluation of the thesis study.

The analysis begins by selecting the most important parts of the interview
and marking them as "code words". This process can follow either a theoreti-
cal approach, where coding is performed for a specific research question, or an
inductive approach, where the research question evolves during the coding pro-
cess (Braun and Clarke, 2006). Then the selected code words will be discussed
and placed in separate themes. Thematic analysis will involve categorizing each
code word into core themes, depending on the amount of information from the
interviews for each core topic, indicating pertinent and crucial research patterns.

This thesis uses an inductive approach in conjunction with coding to iden-
tify themes strongly connected to the interview data. The use of an inductive
method is appropriate since the data is gathered explicitly for this study, in
this case, through interviews. The chosen themes are linked to specific ques-
tions asked to the participants during the interviews, which will be analyzed to
facilitate discussion and draw conclusions.

3.4 Selection of interviewees

For this thesis study, stakeholders from a private company were selected for
the interview using *purposive sampling*, as defined by Denscombe (2010), which
requires the researcher to have prior knowledge of the interviewees. In purpo-
sive sampling, the interviewees are selected who are likely to produce the most
valuable data and relevance to the topic of investigation (Denscombe, 2010).
This selection method will aim to gather the best information from individuals
most likely to have both experience and expertise in ML. To ensure that the
selected interviewees can provide the most authentic information regarding ML implementation in projects, the criteria for stakeholder selection prioritized individuals with sufficient technical knowledge about ML and who work closely with construction project data management. Project leaders and senior individuals were given additional priority as they could provide proper guidance, quality information, and a deeper understanding of the risks involved. For this thesis, the researcher received individual guidance from the industry supervisor, who possessed information regarding the most relevant stakeholders.

3.4.1 Respondent profile

<table>
<thead>
<tr>
<th>Respondent</th>
<th>Job title</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>Head of Business Area</td>
<td>29 min</td>
</tr>
<tr>
<td>R2</td>
<td>BIM Strategist</td>
<td>25 min</td>
</tr>
<tr>
<td>R3</td>
<td>Lead Structural Engineer</td>
<td>25 min</td>
</tr>
<tr>
<td>R4</td>
<td>Head of Digital Transformation</td>
<td>45 min</td>
</tr>
<tr>
<td>R5</td>
<td>Construction Management Consultant</td>
<td>30 min</td>
</tr>
</tbody>
</table>

Table 3.1: Respondent profile

R1, R2, and R3 are familiar with the concepts of ML but do not have knowledge about practical implementation in the sector. R4 has the most knowledge and is familiar with different use cases and implementations in various projects. R5 has past experience working with ML and is involved in the discussion of using ML in the present organization.

3.5 Interview guide

The interview guide will be used to collect data, identify the important information for conducting the result analysis, and determine the purpose of the data collected.
Questions

1. What is your role in your current organization?
2. How familiar are you with data handling in construction projects?
3. How is unstructured data handled in construction projects today?
4. How do construction data differ in different projects?
5. How familiar are you with ML or deep learning?
6. Have you come across the use of AI in your domain?
7. How is text-based construction data handled today?
   a. How can it be handled using ML?
      i. What are the challenges associated with it?
8. How is image-based construction data handled today?
   a. How can it be handled using ML?
      i. What are the challenges associated with it?
9. What do you think are the benefits of incorporating ML in construction projects?
10. In your opinion, what are the main challenges that restrict the use of ML in projects?
11. How would data quality affect the implementation of ML in projects?
12. How would the ownership of data affect the implementation of ML in these projects?
13. What are the risks associated with the implementation of ML in construction projects?
   a. How can the risks be mitigated?
14. How can you ensure that the ML models used in construction projects are accurate and reliable?
15. How can ML create value for project-level analysis?
16. Can you speculate if ML can be used to standardize construction data?
   a. Are you currently exploring or considering investigating ML’s use in projects?
      i. Yes: How would it benefit the project?
      ii. No: Why would it not benefit the project?

Table 3.2: Interview guide

3.6 Ethical consideration

Denscombe (2010) outlines four main ethical principles a study should abide by to set research boundaries and guidelines. The four principles are (1) participant anonymity, (2) voluntary and well-informed participation, (3) refraining from any form of deceit and conducting the research with scientific integrity, and (4) following the legal framework of the country where in which the research is conducted in.

The conducted research study addressed all of these outlined ethical princ-
ciples for all the interview participants, and after approval and consent, the interview continued. This thesis study took ethical aspects seriously, as the company in question may refer to private information during the interview. The anonymity of the participants is maintained throughout the paper, where the participants are addressed using pseudonyms, and the participants contain no obligation to provide personal details to the researcher during the interview. The organization’s name is mentioned in the thesis, which is approved by the organization’s superior and will even be informed in the consent form. Voluntary participation is considered by filling out consent forms before the interview. The consent form informs the research origin, purpose, structure, process, data handling, and the freedom of the participants to terminate the interview at any point. The consent forms are asked to be signed and sent to the researcher before the interview. Considering the EU law, since the interviews are conducted in Sweden, the General Data Protection Regulation (GDPR) law is applied to protect participants’ fundamental rights and freedom to protect personal data. The interview responses in the thesis may be published on an academic public platform such as DiVA, but the responses will be handled ethically considering the GDPR law by informing the participants beforehand and getting approval from the organization. Therefore, the researcher must follow the written considerations to provide ethical security for the participants.

3.7 Alternate strategy and methods

3.7.1 Alternate research strategy

Another suitable alternative research strategy would be conducting surveys to measure social phenomena or trends. Internet surveys can provide comprehensive coverage and yield essential information on the implementation stages of new technology like ML, data management in construction projects, and data usage and storage (Denscombe, 2010). Although internet surveys seem quick and inexpensive to gather results quickly, their reliability is difficult to establish with limited participation. The organization is comparatively small, and a broader approach requires finding knowledgeable participants in ML. Moreover, the survey method would not enable us to obtain a deeper understanding of concerns related to ML implementation, which is crucial for understanding the implications of such implementation in projects of varying sizes. Given the limited availability of resources, conducting surveys is not a feasible research strategy for this thesis study.

3.7.2 Alternate data collection method

Participant observation may be another method of data collection for this study, where the researcher spends time with people or communities to gain an understanding of them. By staying close to the spatial phenomenon being studied, the researcher can observe and learn from their experiences (Laurier, 2010). This
involves working closely with participants, conducting interviews, and seeking knowledge from experts on ML within the organization.

Due to limited access, participant observation is not a viable option for this study. The ideal study participants are based in another city in Sweden, making it difficult for the researcher to be present in their working environment. Additionally, participant observation heavily relies on the researcher as the research instrument, making it challenging to repeat the study for reliability. Finally, this method raises ethical concerns, as participants may feel pressured by being observed. Although some may not agree to participate, they will be required to be present on the same premises as the researcher due to other study participants.

3.7.3 Alternate data analysis method

Another way to analyze the data acquired in this study is through grounded theory which involves data analysis through an elaborate set of coding processes (Walke and Myrick, 2006). Unlike codes used with quantitative data, codes in grounded theory are open to change and improvement as research progresses. Initially, the codes are mostly descriptive, known as open coding (Denscombe, 2010). However, using grounded theory is unsuitable for this study since this method aims to create new theories (Ega, 2002). In this thesis study, open coding can emerge new interpretations from the interviewees, and continuous new interpretations can be extensive and time-consuming. The organization may need to give additional time, which may be problematic for completing the thesis study. On the other hand, the study aims to analyze the risks based on current ML implementation, thus, there may be no requirement for continuous feedback as the research progresses.
Chapter 4

Results

Results This chapter will present the study's results as a qualitative method using semi-structured interviews. The interview results have been transcribed and divided into five different core themes. The themes have been extracted from sub-themes taken from the code words that the interviewees have highlighted. Furthermore, the benefits of using ML will also be discussed, which have been mentioned by the interviewees.

4.1 Identified themes

After conducting a thematic analysis of the interview data, the findings have been analyzed. The five core identified themes from the five interviews will be presented visually through diagrams, followed by explanations incorporating relevant statements from the interview.

4.1.1 Data handling

Figure 4.1: Core theme 1 “Data handling”
Within this section, the focus is on the variety of data formats used and how it is handled among skilled professionals. It deeply discusses several tools used for processing text- and image-based data. Furthermore, it also discusses the complexities involved in data management, including challenges related to traceability, human error, standardization, and insufficient data systems, all influencing the integrity and reliability of construction projects.

Each project in the construction industry has unique deliverables, goals, objectives, and current project state, with different data sources. As stated by (Soibelman et al., 2008), some essential data variables include text and image-based data, which this analysis has focused on.

"Text-based unstructured construction data (like contracts, differences and change of orders) is handled in a file management system” -R1

"Each project is allocated a project number where all the data is stored in the right folder on a SharePoint-based partition, whereas other files, depending on what software is used, were accessing the files stored on the more traditional server-based” -R3

Both R1 and R3 talk about how the files in various projects are stored and the systems used which can be accessible easily amongst the project team. In the case of project-specific files, those are stored depending on what is supposed to be delivered with primary departmental structures. These rules are also applied to image-based construction data captured by cameras, drones, or other imaging devices to provide visual information about the construction project. Although R4 highlights that the handling of files depends on both the company’s size and maturity level, smaller and less mature companies typically manage data in an unstructured way. This involves practices such as sending project-related information via emails to specific individuals, leading to information that is not systematically linked or traceable.

The core difference in construction is the different datasets between different projects. R5 mentioned that every construction project has requirements, specifications, and parameters, resulting in a wide range of data requiring careful management. As a result, dealing with this diverse data calls for specialized knowledge and expertise. Effective and meaningful decision-making in each construction project hinges on the ability of skilled professionals to conduct unique data analyses tailored to the project’s specific needs. R1 refers to this data as unstructured data often handled in project portals or file management systems. Any text-based data in PDFs or similar files are structured into project-specific folders with the project number. R3 supports that statement by describing the collaborative platform used as bellow,

“some files are stored on a SharePoint based partition, whereas other files, depending on what software is using, were accessing the files stored on the more traditional server based” -R3
Image-based data such as 2D and 3D drawings and model outtakes are not widely used. As per R1, all products or construction projects have illustrations that are stored in the same manner in a file structure. R5 mentions that these illustrations were previously printed and stored physically, but now they have been digitized to a certain point. Humans now make these illustrations through software such as Computer-aided design (CAD) and save them as files. With the data, some concerns affect the integrity of construction data. R4 mentions that projects are handled with issue-handling systems, such as JIRA, which are then connected to particular objects or attachments it is related to. However, in many cases, these attachments might be communicated through emails, which leads to difficulty in the traceability and analytics of the data. R5 highlights the issues in construction data, which summarizes that human handling of text- and image-based data leads to human errors in data entry, old information, lack of standardization and digitalization, and incomplete data, which affects the reliability of the project and targets its goals with challenges. R5 further emphasizes that insufficient data management systems further contribute to challenges in ensuring the accuracy and quality of construction data.

4.1.2 Technological advancement

Figure 4.2: Core theme 2 “Technological advancement”

This subsection emphasizes the necessity for technological advancement in managing data within the construction industry. It discusses the significance of data quality, the potential of AI and ML algorithms for improving data handling, and the challenges associated with unstructured data while also addressing the limitations regarding data usage due to project-specific security constraints.

Technological advancement is necessary to improve how data is handled in an organization, which is the core to improving business processes. R2 explains that insufficient interest in producing data leads to organizational drawbacks, whereas, the power of data can be used with an issue management system, which is a more structural way of data handling. Many conclusions can be drawn from the system, such as when adding an image to an issue in the issue management system, the system tries to tag the issue based on the image
type. Although the results were not great for tagging, that was one of the first things implemented in the system, thus, AI or ML algorithms can help plan for such events. R4 explains the lack of project delivery, where, irrespective of the project role, employees require an overview of the project, which is often missing due to unavailable information and sources. R4 emphasized that implementing 3D drawings or 3D models in the new paradigm needs to adapt a new mindset and proper technical parameters, such as AI and ML for project delivery. The construction sector is heavily producing unstructured data, which is dealt with on a general basis. R2 describes that handling this data is difficult as there is data loss due to misspellings and incorrect use and interpretation of the data. ML could potentially help with structuring, for example, text data in a tag system that can filter on keywords. Using such tag systems in the industry will be revolutionary for data quality, but it may affect in a generalized project scope because,

"every project has its own security level, I have to say I’m not sure we can draw upon that on a company level because the data might be restricted to only be used with clients or with a civic project" -R3

Implementation of ML in construction projects depends on the project’s scope, as explained by R3. However, it produces high risk as it restricts data accessibility to specific projects, thereby challenging the utilization of these algorithms in a standardized manner. Consequently, the diverse security measures across projects might hinder smooth integration, thus limiting their overall scope and efficiency.

4.1.3 ML implementation necessities

This subsection underlines the importance of trust, data quality, and training for the successful integration of ML in the construction industry. It emphasizes that improving data quality through standardized naming conventions is necessary for enhancing ML’s usability and effectiveness in construction projects.
In the construction industry, implementing advanced ML algorithms highly depends on the organization’s maturity, which both R2 and R5 have highlighted. R5 further mentions,

“there should be a level of robustness and reliability of the ML system that is being deployed” - R5

R2 elaborates that the immaturity of people in the construction industry would make them hesitant to use ML, whereas, with some training, it will be easier for such implementation. The construction industry is risk-aware, and trust is the key to enabling ML use in projects. There is a lot of data, for example, safety data, which can be trained to efficiently reduce the risks, recognize them, and suggest ways to eliminate them. Using ML in this manner is both efficient and quick, thus, having access to such data is crucial as it enhances robustness and enables generalization of new data. The data produced can keep on training itself, enhancing the performance and usability of construction projects.

R3 explains that data quality has a huge effect as checking all the data produced from construction into ML models is well used to make accurate decisions. In construction, there is data from various projects that are not well-established, therefore, it is difficult to distinguish which part of the data is quality-assured. It is necessary to carefully assess and validate the data to ascertain its accuracy and credibility, leading to more reliable outcomes when training ML models and making informed decisions in practical applications. R1 further adds that quality assurance is the key to maintaining order and establishing solid decisions on the right foundations. By accumulating data from construction projects, computer-based decision-making can streamline the process for decision-makers. Project managers can infuse their expertise into the data pool, and as the dataset expands, the resulting decisions will gain depth and precision.

In furthering the enhancement of data quality, another critical aspect highlighted by R4, which could be seen as a challenge, pertains to the varying naming conventions and data formats in use. ML and AI have the potential to contribute significantly by recognizing distinct types of objects and effectively mapping them or translating one into another. An illustrative example is the generation of BIM models from 2D drawings, where the technology can transform data found within a PDF drawing into a cohesive 3D model. Adopting standardized naming conventions is pivotal in promoting data consistency and improving data quality within the construction industry.
4.1.4 Benefits with ML implementation

This section explains how ML implementation presents a unique approach to efficiently address everyday inquiries, potentially easing workload burdens on product managers and enhancing operational efficiency. ML can analyze project statistics, identify underlying issues, and suggest solutions while helping analyze root causes by looking into unstructured data. Additionally, ML enables examining safety data to identify risks, supports standardization through 3D model adaptation, and facilitates decision-making by comparing alternative solutions from different projects.

Utilizing ML or any generative AI system represents a distinctive approach, as these models promise to address everyday inquiries efficiently. This could potentially remove a significant workload burden on product managers. Introducing such systems to handle routine questions has substantial potential for enhancing operational efficiency (R2). Another advantage explained by R4 states that ML can be used to look at the statistics of projects relative to other projects.

"...you can use machine learning, for example, to look at the statistics of projects relative to other projects..." -R4

This project comparison helps identify potential issues and underlying causes within project management and among various stakeholders. Although ML can also be integrated into dashboards, these dashboards typically do not offer insights into the fundamental reasons behind issues. Therefore, a more in-depth root cause analysis can be conducted, potentially leveraging the analysis of unstructured data within the issues’ content (R4). ML can also enhance project analysis by looking into safety data to identify the most common types of risks and suggest ways of fixing them or making people aware of these problems. It can also assist in evaluating on-site hazards by automatically detecting safety-related observations and providing warnings regarding potential risks (R4).

Furthermore, ML supports standardization and the development of standardized solutions. For instance, if there is access to 3D model data of a building, it becomes possible to adapt these models to specific project requirements. Examining data from various projects and assessing how alternative solutions
applied in different circumstances could be relevant to the current project for a more efficient decision-making process (R1). R1 further highlights that ML can facilitate the integration of different aspects within construction projects, allowing them to concentrate on various connections from 3D models. By utilizing a dataset encompassing these connections across different project types and various products, users can get solutions tailored to their needs. This enables them to search the database and find suitable solutions based on the client’s requirements (R1).

### 4.1.5 Challenges with ML implementation

![Core theme 5 “Challenges with ML implementation”](image)

The respondents mentioned four challenges with implementing ML in construction projects: (1) data fragmentation, (2) security restrictions and data leakage, (3) trust and culture, and (4) cost and data utilization. R2 discusses data fragmentation, which starts with proper input and is visualized with the data produced by various project accountables, such as contractors and organizations. This data is mostly valuable to the project owners or clients who benefit more from future data quality. However, the client push acts as a higher effect because of the fragmentation of different projects for the infrastructure. R5 highlights,

> “Construction projects get data from various stakeholders with their own data formats and quality levels. ML models require consistent and high-quality data to produce accurate results” -R5

This explains that different data produced by the data owner makes the data accustomed to its requirement, which generates poor quality data that can be considered biased or unreliable. Project fragmentation is often caused by a lack of standardized processes and data formats in the construction industry, which causes a higher impact on the different projects. In order to implement ML models, the data is required to operate in a standardized manner.

Another challenge with the ML implementation mentioned by R3 is security
restrictions that are entitled to every project with its security level. This means the produced data may be restricted to be used only within the project team and its clients. Although this limits drawing bigger conclusions, it prohibits a more profound risk that the construction industry may face, which is data leakage. As each project has unique requirements, the restrictions are higher, and the possibility of openly discussing different projects is not permissible. Such projects cannot be included in a database that the involved project parties do not solely control. R3 further mentioned about the liability of using ML,

“if we are to build something that’s a little better structure based on artificial intelligence, then the question comes up where do the liability lie? I mean, if the building collapses, who is responsible for that collapse? Could you really trust this artificial intelligence...” -R3

This summarizes the importance of the reliability of the data as it can introduce a force of security for the ML models. R5 emphasizes the significant impact of reliable data on an organization’s people and culture when implementing ML models. It underscores that trust in the data quality is fundamental, as the absence of such trust can erode confidence in the machine models, as previously highlighted by R2.

R1 further underscores a critical concern,

“you have to be able to trust the data. And I think that’s kind of one of the challenges for the construction industry...” -R1

Trust is a primary challenge for individuals to misuse data, which may manifest as unwarranted assumptions. When data-driven results come from a computer, there can be a tendency to accept them unquestioningly, assuming that data lends accuracy. The essence of this concern lies within the domain of culture and individual behavior, necessitating a collective shift in attitudes, especially among clients. Clients, who may vary from product to product, must adapt their perspectives collectively, and this cultural transformation is pivotal to facilitating innovative practices, as the implementation of ML often encounters substantial resistance and challenges. On the other hand, stakeholders must also have confidence that these models provide well-informed and impartial predictions or suggestions. An organizational culture that embraces the importance of refining ML models with essential data can effectively utilize these mechanisms, as stated by R5.

Another challenge of handling data is mentioned by R1, where we know that the construction industry is not familiar with handling data using ML, therefore, the industry does not think of the possibilities. In contrast, individuals in organizations who see the possibilities do not know how to explain them to others. For example, for the project clients, why should they use ML when the cost of implementation is so high? A client might have five new projects a year and have a lot of different properties from history, so the data produced can be
little data to be used efficiently. There is no certainty if the data is valuable or not, thus, the industry is far off from software development where the gap is high for us to utilize it, which is a risk as we cannot depend on the data quality. The factor stands for clients with many properties or a lot of old data or unstructured data already, which will be challenging to get running for ML models, which are always highly data dependent (R1).
Chapter 5

Risk identification

The gathered qualitative data will be analyzed further through risk identification, which will be evaluated through a risk matrix. It will also discuss the benefits of implementation to understand the construction operations.

5.1 Risk identification

A key component of risk identification will help the organization to identify if an ML implementation will improve data handling, project management, and operations in construction projects. Based on the five interviewees, nine risks have been identified, which have been referred to according to the risk structure in construction as per Zavadskas et al. (2010). These risks have been differentiated between internal, external, and project risk and which category they belong to. The following table will map out those specifications for each objective and determine if ML implementation benefits their projects.

<table>
<thead>
<tr>
<th>Risk ID</th>
<th>Description</th>
<th>Risk type</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Insufficient knowledge of data (R1)</td>
<td>Internal</td>
<td>Information</td>
</tr>
<tr>
<td>2</td>
<td>Limited trust in data and ML implementation (R1, R2)</td>
<td>Internal and External</td>
<td>External: Information, Internal: Social, External: Economic, Project: Cost</td>
</tr>
<tr>
<td>3</td>
<td>ML implementation is costly (R1)</td>
<td>External and Project</td>
<td>Project</td>
</tr>
<tr>
<td>4</td>
<td>Gap in software development (R1)</td>
<td>Project</td>
<td>Technological</td>
</tr>
<tr>
<td>5</td>
<td>Data quality varies from project to project (R1)</td>
<td>Project</td>
<td>Quality</td>
</tr>
<tr>
<td>6</td>
<td>New implementation is an obstacle to adapt change (R1, R2)</td>
<td>External</td>
<td>Social</td>
</tr>
<tr>
<td>7</td>
<td>Higher project fragmentation (R2)</td>
<td>Internal</td>
<td>Documentation</td>
</tr>
<tr>
<td>8</td>
<td>Data leakage (R2, R5)</td>
<td>Project</td>
<td>Quality and Technological</td>
</tr>
</tbody>
</table>

Table 5.1: Risk identification
5.1.1 Risk evaluation

The risk evaluation connects the risks identified with its impact and likelihood. Within this process, the fundamental impact and likelihood play a vital role in evaluating and controlling potential risks. These concepts aid in understanding the severity of the risk, thereby facilitating prioritization and management of these risks.

<table>
<thead>
<tr>
<th>Risk ID</th>
<th>Impact</th>
<th>Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High</td>
<td>Likely</td>
</tr>
<tr>
<td>2</td>
<td>Medium</td>
<td>Likely</td>
</tr>
<tr>
<td>3</td>
<td>High</td>
<td>Very likely</td>
</tr>
<tr>
<td>4</td>
<td>High</td>
<td>Very likely</td>
</tr>
<tr>
<td>5</td>
<td>Medium</td>
<td>Likely</td>
</tr>
<tr>
<td>6</td>
<td>Low</td>
<td>Likely</td>
</tr>
<tr>
<td>7</td>
<td>Medium</td>
<td>Very likely</td>
</tr>
<tr>
<td>8</td>
<td>High</td>
<td>Likely</td>
</tr>
<tr>
<td>9</td>
<td>Medium</td>
<td>Likely</td>
</tr>
</tbody>
</table>

Table 5.2: Risk evaluation

5.1.2 Risk matrix

The 4T’s of hazard of risk management signifies risk by classifying them between tolerate, treat, transfer, and terminate. Hopkin (2017) explains, (1) Transfer, presents the risk to another party, (2) Tolerate, explains the impact of the risk, (3) Terminate, are the activities that are generating the risk, and (4) Treat, helps to reduce the impact or exposure of the risk.

Figure 5.1: Risk matrix and the 4Ts of hazard management
Chapter 6

Discussion and conclusion

This chapter begins with a summary of the study and then discusses the research objectives based on the thesis findings. Furthermore, the study’s limitations and ethical and societal consequences are discussed.

6.1 Discussion

Implementing ML in construction is an under-researched area, although, as Van and Toan (2021) mentioned, ML can significantly improve the management of such projects in different areas, from cost to prediction. On the other hand, Regona et al. (2022) highlighted several real-world challenges in implementing ML within the construction sector. Firstly, many organizations and project stakeholders within the construction sector lack knowledge when implementing new technological solutions, making ML implementation solutions potentially complex for the stakeholders to understand. This industry is running behind in incorporating ML or AI mostly because of the lack of structured data. This leads to another obstacle due to the subjective nature of data collected from various construction projects, making it challenging to utilize the data for training AI models. Accordingly, trained models may exhibit incompleteness and data consistency due to the fragmented nature of the industry. Even when leveraging large datasets, the reliability and security of the data are challenged since each dataset is project-specific based on project structure, objectives, and requirements. This fragmentation across projects also contributes to a lack of standardization within the construction sector, making ML implementations risky and costly, particularly if they fail to provide valuable insights. Furthermore, adopting such technologically advanced solutions necessitates significant financial investment for maintaining and continuously training AI models to ensure they remain up-to-date and capable of producing accurate results.

The investigation done in this study was based on the limitations and problems that could happen when making project decisions based on ML models.
With the help of the interviews and thematic analysis, this thesis could effectively find the most significant concerns with ML implementation and how they can be tackled in the future. The thematic analysis’s various themes and the mentioned challenges and benefits underline the most important aspect to answer the research objectives. Highlighting the benefits of ML implementation, this thesis even discussed the value creation in operations compared to a project-level analysis.

**RQ1: What are the main risks and challenges of implementing ML in construction projects, and how do they impact construction operations?**

There are several risks and challenges when using ML in construction projects. One big issue is that each construction project is customized for a specific client, so the data available is inconsistent. This causes problems when we want to use the same model on different projects because it cannot be relied upon to produce the same results. Additionally, some projects do not generate sufficient data, making it challenging to train and validate ML models properly, and their decisions may not be trustworthy. This can be risky for construction operations that rely on data-driven decisions. Educating the organization about ML and building trust in data quality is essential to reduce these risks. Resistance to new ML solutions can slow their adoption, and both clients and stakeholders should change their attitudes and help generate better data. Even limited data sharing can limit the ability to make broader conclusions, which hampers the potential benefits of ML. Addressing these challenges to successfully integrate ML into construction operations and avoid problems when making data-driven decisions on construction projects and sites is essential.

**RQ2: How can ML contribute to value creation in operations compared to a project-level analysis?**

To ensure sound decision-making foundations, it is important for projects to undergo quality assurance processes. When a predefined ML model analyzes existing data and learns from past mistakes to offer more fruitful suggestions, it simplifies decision-making. Implementing ML in construction operations translates to enhanced predictive maintenance capabilities, as ML models can predict risks associated with various aspects, including safety concerns and environmental factors. By forecasting potential risks, organizations can take necessary actions to mitigate these incidents. ML can efficiently allocate resources such as labor, materials, and machines in operational contexts, optimizing their usage. ML can also uncover patterns and trends in operational and project-level data, enabling organizations to refine their processes and make data-driven choices. Moreover, ML’s ability to self-improve through iterative data analysis enhances its value in operational and project-level analysis.

Regarding project-level analysis, ML can improve image processing in construction projects involving 3D or 2D images. By integrating image recognition and data analytics, ML can enhance quality control by identifying image components and applying the knowledge to future projects. The same concept applies
to text-based construction data, where ML can be trained to facilitate building maintenance operations. ML models can also provide accurate project cost estimates by analyzing historical data and adjusting for project-specific variables, helping projects avoid budget overruns and ensuring profitability.

RQ3: How is ML beneficial for the future of projects in the construction sector?

Utilizing advancements in technology to manage projects and streamline the handling of construction data efficiently holds great potential, particularly in expediting project decision-making processes. The construction industry produces a substantial amount of unused data, which can be analyzed through ML and AI to provide tailored solutions based on the unique requirements of each project. Initially, the focus can be on processing text and image-based data to predict insights related to construction sites, various contracts, specifications, construction models, and images. Given the substantial volume of project data, it can be organized within a more robust data framework, allowing cross-project analysis. ML can also benefit by comparing different projects, helping identify underlying issues for future projects, thereby improving knowledge exchange and cost for each project. Forecasting and prediction play a crucial role in ML, as they can effectively identify potential hazards, thereby improving working conditions at construction sites. While the potential benefits of ML in the construction sector are substantial, successful integration requires investments in data collection, improving data quality, and the implementation itself. Furthermore, enabling collaboration between project stakeholders and technology experts is essential to maximize the impact of ML on construction projects.

6.2 Conclusion

In conclusion, this thesis has discussed the risks and challenges of integrating ML into construction projects, highlighting data inconsistency. When we start new construction projects, we are building new data, thus, we need to ensure the data is reliable. Addressing this challenge is crucial to ensure the reliability of data-driven decision-making in construction operations. Furthermore, this research has analyzed the potential for ML to contribute to value creation in construction operations and project-level analysis. ML’s ability to enhance predictive maintenance and data-driven processes benefits the industry. However, due to insufficient data to run ML models, the initial approach is that instead of using ML for an entire project, it can be implemented for specific parts or datasets and compared to other datasets in different projects to make a more significant analysis which will act as an initial step to put these ML models into practice. There still needs to be a significant link between ML and its practical application in real-world scenarios, even when the ML system shows reliability comparable to that of a human. ML has not yet matured to a level where researchers from various areas can easily apply ML to their specific problems. While ML offers valuable strategies for tackling complex problems, current re-
search frequently lacks a direct linkage to real-world issues, highlighting the need for a more integrated and problem-centric approach.

Moreover, stakeholders of different projects need to have trust in such implementation to gather more reliable data. It requires more knowledge development and expertise, and such implementation will be fruitful for future project development. Thus, appropriate data must be utilized without exposing sensitive project data. Due to project individuality, the construction industry requires more research as it is underdeveloped compared to other technologically advanced fields. Although, ML holds great potential for future projects in the construction sector as its predictive capabilities and with the right approach and commitment, ML can revolutionize the construction industry, paving the way for more efficient, data-driven construction projects. This study has the potential to be used as a basis for future research that could help more organizations in using ML in certain areas of construction projects, which produces more structured data that can be used fruitfully. This could result in organizations having a better understanding and knowledge transfer for a more efficient implementation.

6.3 Limitations

Amongst other limitations, an important limitation was that, on many occasions, the respondents did not have enough knowledge to provide an extensive analysis of ML implementation. Repetitive and follow-up questions have been asked variably to gather more project information. However, limited knowledge might have affected the results and the reliability of this thesis’s results. Another limitation that may have affected the results is the number of participants who could participate, affecting the credibility of the research. Participants with more industrial knowledge of ML implementations in construction could have helped by helping with a more in-depth analysis of the risks and future prospects of ML in this area. Another limitation was confidentiality, as information may not have been shared as it contained private and individual construction projects. As each project concerns various stakeholders, the answers may have been generalized for a broader group with minimal project information. In addition, since all the respondents were based in different organizations, they are working with different solutions and knowledge, contributing to increased credibility of the research as they have highlighted different situations where ML can work or fail. On the other hand, as the respondents were not exclusively familiar with each other, it did not influence the answers, which strengthens the reliability of the study.

6.4 Ethical and societal consequences

The result is not considered to have any ethical or social consequences as it only discusses the risks and future research prospects to develop the construction
section in terms of more digitalization, automation, and using more advanced
technological solutions, which cannot have any negative impact on other re-
search or organizations or on the individuals who participated as interviewees.
However, one consequence may be that more research in this area may be limited
as it is costly and such implementation is underdeveloped.
Chapter 7

Future research

In this last section recommendations are made for future research based on the impact of the findings.

The discussions in this thesis are only suggestions on how organizations can enter the field of ML to standardize their data effectively in the sector. Depending on its focus areas, this approach can be personalized to align with specific projects. In terms of future research, there are opportunities to explore deeper into the implementation of predictive models for construction projects. ML has the potential to enhance various data types, including images, text, project costs, and the identification of potential risks. Achieving this involves the development of more advanced algorithms that incorporate historical data and automate decision-making within construction processes. These models can also enhance safety on construction sites by providing real-time data and identifying potential hazards, simplifying project operations and management. However, with more collection and sharing of data across different projects, research into data security and privacy becomes important. Therefore, establishing robust security measures and ensuring compliance with data protection regulations will be crucial. Future research can include creating training programs and educational materials to upskill construction professionals in AI and ML technologies. Subsequent research in this area can pave the way for adopting ML in construction, ultimately leading to more efficient, cost-effective, and sustainable infrastructure development. Collaboration between academia and the construction industry will propel this field forward.
Chapter 8

Bibliography


Chapter 9

Consent form
INFORMED CONSENT FORM

Risk analysis of implementing Machine Learning implementation in construction projects

Background and purpose
The study is carried out for a master’s thesis at the Department of Computer- and Systems Science (CSS), Stockholm University. The research goal of this thesis is to provide insights into the current state of Machine Learning (ML) implementation in construction projects and identify the challenges and opportunities associated with its use.

How does the study work?
The study is carried out in reference to a master’s thesis in Strategic Information Systems Management. The thesis analyzes challenges, risks, and success factors using ML to standardize construction data. The study also explores the benefits of such a study in enhancing the comprehension of ML’s application in various construction projects. With the help of the interviews, the researcher can determine how ML could be applied to unstructured data and the future requirements for successfully integrating ML into such projects. The interview protocol comprises a series of open-ended questions to encourage interviewees to share their experiences and insights on the topic. The interview consists of approximately 10 questions which will be recorded with the interviewee’s approval. The interview is estimated to take place with a time frame of 45 minutes.

Handling of data
The data will be handled by various stakeholders at OSV and can therefore be published on various platforms as a learning tool. The data is stored in a way that prevents it from being lost, becoming corrupt, and intrusion by unauthorized persons. Pseudonymization has been used to anonymize and protect the identity of the interviewees.

Volunteering
In accordance with the GDPR law, participation in the study is voluntary and can be discontinued at any time without any further explanation.

Responsible
The study is carried out by Aki Roy
Telefon: + 46 (0) 73-513 38 70

I confirm that:
I have received the nature of the research study in written and oral
I give my consent to participate in the study and am aware that my participation is voluntary
I am aware that I can end my participation at any time without an explanation
I allow the information shared and the data collected to be stored and handled electronically by the thesis supervisors

Name
First Name

Last Name

44
Chapter 10

Reflection Document

The initial thesis topic was suggested by an industrial organization interested in integrating ML into their data management practices. As a student with a background in data management, this subject presented both a challenge and a novel field for exploration. Despite being unfamiliar, the primary aim was to deepen my understanding of the application of ML in construction, with the study intended to provide a comprehensive overview of the concept. Throughout the essay, I delved into the fundamentals of ML and reviewed prior research in the field, establishing crucial groundwork for constructing the research methodology. Upon careful examination of conducted interviews, it became apparent that a thorough risk analysis was necessary. This shifted the thesis focus from data management to risk analysis associated with employing ML in construction, aligning the study more closely with real-world examples. Following the identification of the research objective, it became evident that the obtained results were insufficient for the full implementation of ML in a project, as it could potentially give rise to other challenges that needed to be addressed first. Consequently, the research objective was adjusted to align with the insights derived from the interviews.

The thesis plan has been conducted in iterations, involving discussions with my supervisor to identify relevant data and literature in the construction domain. I have had bi-weekly meetings with my supervisor, who provided insights into the current industry processes and helped identify relevant industry papers to support the research objectives. After setting research objectives, stakeholders essential to the study were identified for interviews. Interviews were scheduled with mutual consideration and coordination based on availability. An area for improvement in this study was the inclusion of individuals in a risk analysis workshop. Organizing a group workshop had the potential for more comprehensive answers, feedback, and discussions about the study’s results. However, arranging a single meeting with all respondents and stakeholders was challenging due to time constraints.

The study has followed the instructions for writing a thesis for a degree project using relevant scientific methods and literature consistent with the re-
search objective. This thesis study is aligned with my education in Strategic Information Systems Management, where I acquired skills in writing scientific papers through method courses such as Scientific Communication and Research Methodology (FMVEK). Additionally, data-related courses, such as, Data Warehousing (IS5) and Open and Big Data Management (OBIDAM) taught me the fundamentals of ML and AI. The Risk Analysis course (RIMA) was crucial in teaching me how to conduct assessments in various domains. Consequently, the subject of the study is something I can work on in the future as it can help organizations understand how they can implement the concepts of ML in construction projects.

The study has presented valuable results and well-founded discussions, serving as an independently conducted academic research that will contribute to my ability to undertake independent work in my professional life. A successful thesis demonstrates my competence in conducting diverse studies and my awareness of the various risks and components of system implementation. It also highlights my proficiency in communicating, conducting interviews, and improving my research and analytical skills. Furthermore, the thesis has deepened my technological knowledge of ML and exposed me to its application in construction. Engaging in research for a master’s thesis leads to a mindset of continuous learning, emphasizing the importance of staying updated with technologies. Despite the study’s time constraints impacting the results due to the missing workshop, I have demonstrated structure, awareness of challenges, and effective collaboration with the organization through well-established communication, contributing to an insightful analysis that addresses the study’s objectives.