A Framework for Enhancing Contract-Related Documentation in Construction

by

Parinaz Jafari

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Construction Engineering and Management

Department of Civil and Environmental Engineering
University of Alberta

Abstract

The management of project documentation involves processing a large amount of important information embedded in different contract and project specification documents. Although contract-related documentation is critical for effective information flow and—in turn—successful project management, it remains a relatively underexplored area of construction management research and practice. The few studies that have explored document processes in construction have limited their focus on the development and improvement of various document management systems. These improvements, however, have failed to achieve the anticipated performance in construction. Documentation requirements remain scattered haphazardly throughout project contracts, complicating their identification and management by practitioners. Moreover, documentation processes are often overlooked and mismanaged, lack efficient planning, and are prone to variability, ultimately resulting in time and cost inefficiency. Structured methods capable of addressing the underlying problems and limitations of contract-related documentation in construction, however, have yet to be developed.

This thesis is proposing a two-phase framework designed to enhance documentation, communication, and sharing practices in both the planning and execution and control phases of construction projects. For the planning phase, a method capable of automating a portion of the administrative process and enhancing decision-support for administrative resource planning is proposed. Here, a natural language processing approach capable of automatically extracting documentation requirements embedded in contract documents was developed. Then, a Monte-

Carlo simulation model was created and used to predict the overhead costs and durations associated with completing contract-related documentation. Application of the planning phase portion of the framework is anticipated to improve estimation and planning of administrative resources, while also enhancing the ability of practitioners to negotiate for the reduction of redundant or irrelevant contract requirements, thereby improving value to all stakeholders.

During the execution and control phase of documentation processes, Lean approaches and network studies are used to enhance the overall performance of documentation processes in construction projects. First, a structured procedure for applying Lean construction principles to enhance and support document management processes through the reduction of hidden waste (such as non-value adding activities) is proposed. In this procedure, value stream mapping is integrated with simulation modeling to quantitatively assess the performance of the documentation process, to identify potential improvements to the current process, and to quantitatively predict the impact of proposed improvements on future project performance. Then, social network analysis is employed to measure and analyze communication of project participants in the documentation process network. Application of the execution and control portion of the framework is expected to reduce waste, rework, omissions, and errors, in turn increasing profit and value for both contractors and clients.

The feasibility and functionality of the proposed framework was validated using practical case studies, the results of which have also provided valuable information for practitioners. Altogether, this research has developed a procedure that can facilitate (1) the extraction of documentation requirements; (2) the forecasting of process time and cost uncertainty measurements; (3) the elimination of excess production and document processing to increase transparency and reduce waste within the administrative process; and (4) the discovery and quantification of documentation process networks for improved efficiency.

Preface

This thesis is an original work by Parinaz Jafari. This thesis is organized in a paper-based format.

A version of Chapter 2 has been published as Jafari, P., Al Hattab, M., Mohamed, E., and AbouRizk, S. (2021). "Automated Extraction and Time-Cost Prediction of Contractual Reporting Requirements in Construction Using Natural Language Processing and Simulation." in *Applied Sciences*, 11(13), 6188, and has been reprinted with permission from MDPI. Dr. Al Hattab was involved in data curation, visualization, and manuscript review, and editing. Mr. Mohamed was involved in data curation, manuscript review, and editing. Dr. AbouRizk was the supervisory authority and was involved in conceptualization, funding acquisition, project administration, manuscript review, and editing.

A version of Chapter 3 was submitted for publication as Jafari, P., Mohamed, E., Hamzeh, F., and AbouRizk, S. "Integrating Value Stream Mapping and Discrete-Event Simulation to Improve Administrative Processes Within the Construction Industry." to the *Journal of Engineering, Construction and Architectural Management* on July 30, 2021. Mr. Mohamed was involved in conceptualization, manuscript review, and editing. Dr. Hamzeh was involved in conceptualization, manuscript review, and editing. Dr. AbouRizk was the supervisory authority and was involved in funding acquisition, project administration, and manuscript review.

A version of Chapter 4 has been published as Jafari, P., Mohamed, E., Lee, S., and AbouRizk, S. (2020). "Social network analysis of change management processes for communication assessment." *Automation in Construction*, 118, 103292, and has been reprinted with permission from Elsevier. Mr. Mohamed and Dr. Lee were involved in conceptualization, manuscript review, and editing. Dr. AbouRizk was the supervisory authority and was involved in conceptualization, funding acquisition, project administration, manuscript review, and editing.

The research project, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board, "Social Network-Based Analysis of Change-Management Processes for Collaboration Assessment," Pro00088819, approved on March 8, 2019.

Acknowledgments

This thesis is a product of the work, guidance, and love from great people that have helped me during the journey of my Ph.D. program.

First and foremost, I take immense pleasure to express my sincere and deep sense of gratitude to my supervisor, Prof. Simaan AbouRizk, for his sustained enthusiasm, creative suggestions, motivation, and exemplary guidance and support throughout the course of my doctoral research. He had a paramount influence on many aspects of my life and built a solid foundation for my educational and professional career.

I am also grateful to my advisory committee, Dr. Farook Hamzeh and Dr. Ahmed Hammad, for reviewing my dissertation and providing their invaluable feedback. I would also like to thank my examining committee, Dr. Irtishad Ahmad, Dr. Yashar Pourrahimian, and Dr. Zhehui Jin, for their valuable insights.

I would like to express my warm thanks to Dr. Catherine Pretzlaw, Maria Al-Hussein, Brenda Penner, Steve Hague, and all the friends that I made during my PhD. I am really grateful for having you in my journey. It was a pleasure to know you all.

Most importantly, I would like to thank my parents and my sisters for their dedication, boundless love, and never-ending support, which enhanced my ambition throughout my educational and personal life.

Last but not least, I would like to express my deepest gratitude to my beloved husband for his boundless patience, support, and love. Your energy and inspiration motivated every step I took along this journey. I would have never been able to finish this PhD without you.

Contents

List of Tables

List of Figures

List of Abbreviations

NLP	Natural Language Processing
ML	Machine Learning
AI	Artificial Intelligence
OCR	Optical Character Recognition
TF	Term Frequency
IDF	Inverse Document Frequency
NB	Naïve Bayes
LR	Logistic Regression
RF	Random Forest
XGBOOST	Extreme Gradient Boosting
VSM	Value Stream Mapping
DES	Discrete Event Simulation
VA	Value-Adding
NW	Necessary Waste
PW	Pure Waste
PM	Project Manager
PCT	Project Control

S	Superintendent
PA	Project Accountant
PC	Project Coordinator
SCH	Scheduler
DSR	Design Science Research
ERP	Enterprise Resource Planning
SNA	Social Network Analysis

Chapter 1
Introduction

1.1 Background and Problem Statement

The management of project documentation involves processing a large amount of important information embedded in different contract and project specification documents. Numerous types of contract-related documents are generated in construction including, but not limited to, contract documents, project specifications, progress reports, quality reports, environmental reports, safety reports, and changer order documents (Al Qady and Kandil, 2010). Indeed, it has been estimated that a single-structure construction project can generate over 10,000 documents (Turk *et al.*, 1994).

Managing construction projects requires controlling the running schedule and cost to ensure they adhere to planned baselines. To accomplish this, project managers representing each stakeholder request periodical reports that represent the status of the project. Many of the reporting documents that must be prepared by the contractor and submitted to the client are mandated by the construction contract and project specifications. Reports are used to predict any potential problems and to consider implementing necessary measures capable of maintaining the project course, budget, and the client's expectations in a timely manner. The vast majority of project information is created, stored, and shared between stakeholders using construction reports, highlighting the importance of document management and communication systems within construction projects (Al Qady and Kandil, 2013).

The successful management of both construction projects and competent decision-making requires effective and efficient management of contract-related reporting documentation. Although critical for effective information flow and, in turn, successful project management, contract-related reporting documentation remains a relatively underexplored area of construction management research and practice. The few studies that have explored document processes in construction have limited their focus on the development and improvement of various document management systems (Das *et al.*, 2020; Macías-Jiménez *et al.*, 2019). These improvements, however, have failed to achieve the anticipated performance in construction (Al Qady and Kandil, 2010; Fernando *et al.*, 2019).

It has been estimated that 60 to 80 percent of all costs associated with meeting client demands and achieving satisfaction are related to administrative functions (Monteiro *et al.*, 2017; Tapping and Shuker, 2003). However, research focusing on other topics beyond document management systems, such as administrative works and soft management of the reporting documentation process, has been side-lined and under-researched (Belayutham *et al.*, 2016). A considerable potential for improvement, therefore, lies in the exploration of such topics. Notable benefits can be achieved by focusing on enhancing knowledge extraction techniques, documentation process workflows, and communication of the reporting process. Any improvement in contract-related reporting processes is likely to impact the performance of the entire construction organization, as essential components of all projects (e.g., construction reports, procurement, payment) flow through these processes (Alves *et al.*, 2016).

When analyzing needs, requirements, and administrative works related to reporting documentation in construction, the content of the contract documents are first reviewed, analyzed and interpreted during the planning and pre-construction phase. However, the increase in size and complexity of construction projects have resulted in a corresponding increase in complexity of construction contracts and information requirements. Reporting documentation requirements remain scattered haphazardly throughout project contracts, complicating their identification and management by practitioners. In current practice, due to a lack of suitable methods and the tremendous effort required to review and manually extract the requirements (Caldas *et al.*, 2002), teams often do not plan and budget according to the specific requirements of a project. Instead, teams simply consider the documentation effort as a portion of the indirect cost of a project. An efficient documentation plan, therefore, is not established before execution begins.

In practice, a lack of pre-construction documentation planning has resulted in documentation processes that are mismanaged, inefficiently planned, and prone to variability—ultimately resulting in wasted time and increased costs. Assumed activity durations are often unrealistic, underestimated, and not representative of the actual time spent on project reporting. During project execution, project teams execute multiple activities to collect and analyze the required data, prepare and circulate the document, review and approve the document, and many other intermediate steps until required documents are submitted to the client and approved (if necessary). These repetitive processes often involve multiple specialists performing wasteful activities, with no indicators to measure process performance. Hines and Taylor (Hines and Taylor, 2000) stated that only 1% of the effort exerted in administration and information management processes is value-adding, with approximately 99% of the effort classified as waste.

Despite the time and effort expended to complete administrative tasks and the impact of administrative processes on overall project performance, inefficient administrative processes are rarely seen as a significant factor influencing production. As such, methods for improving the efficiency of construction administrative processes remain relatively unexplored (Belayutham *et al.*, 2016), and quantitative measurement of documentation process durations and costs has not been investigated in the literature. Improvements in administrative processes, therefore, remains an unexplored area of construction that could result in notable positive impacts on overall time and cost, elimination of rework, productivity increases, and improved product quality (Sastre *et al.*, 2018).

Moreover, strong communication and analysis are essential to fulfil the documentation requirements of any contract and to avoid delaying construction work on site (Garrett and Lee, 2010). In a construction documentation process, a network of project members work together and enter into various communication arrangements to create value and achieve project goals. Project members must exchange large amounts of information about activities, processes, and decisions used to deliver the project. In documentation processes, communication plays an important role in project success. If a contractor submits reports and related documents late, incompletely, or with deficiencies, it can negatively impact the construction schedule—particularly when the start of a construction phase is dependent on such approvals (Chin, 2009).

Structured methods capable of addressing the underlying problems and limitations of contract-related documentation in construction, however, have yet to be developed. Without a guiding framework, the documentation process can seem complex and ambiguous to clients and contractors. Four contract-related documentation barriers that have yet to be addressed in literature have been identified:

1. It is a challenge to identify documentation requirements from thousands of pieces of unstructured data typical of most contract documents. Manually extracting the required contract-related documentation is impractical due to the large number of documents and unique project specifications—especially during the planning phase when time is limited.

2. Documentation processes lack transparency. Although a large administrative burden on contractors, the time and resources needed to complete reporting requirements are often unknown and unaccounted for during the preliminary planning stages of a project. The quantitative measurement of reporting process durations and costs must be estimated using a combination of historical data and expert knowledge.

3. The documentation process, including generating, managing, analyzing, and communicating documents, is associated with a high amount of waste and non-value-adding activities. Moreover, indicators capable of assisting performance of the administrative process are lacking.

4. Project members must efficiently communicate and circulate documents in a timely manner. In the current process, however, there is no control over or means of assessing the communication between team members. Methods for improving access to the information contained in construction contract documents and streamlining their processes have not

yet become an essential component of construction document management.

The primary challenges of contract-related documentation, such as lack of efficient planning, manual review and analysis of contract documents, low execution efficiency, and lack of criteria and indicators to quantify, assess, and control the performance and communication within the process, were identified as problems with practical relevance and research potential. Here, four questions were posed to address the four barriers limiting the contract-related documentation process:

1. How can contract-related documentation requirements for new projects be extracted more efficiently?
2. How can the time and cost needed to complete required documents be estimated more accurately using available information?
3. How can current documentation processes be assessed, quantified, and streamlined?
4. How can documentation process networks be assessed, quantified, and streamlined?

1.2 Research Objectives

Construction reports are essential for successful information flow between contractors and clients—both on and off the site. Most problems in construction are attributed to communication and reporting problems. Links between the various stakeholders must be built on robust communication and reporting systems that allow the acquisition of precise, simple, and accurate information. Lack in communication and reporting system can result in the circulation of misleading information, which may consequently lead to poor decision-making.

The overall goal of this research was to develop a framework capable of addressing each of the aforementioned research questions to enhance reporting documentation, communication, and sharing practices in both the planning and the execution and control phases of construction projects. This research designed a two-phased framework to achieve the following objectives:

Objective 1: Automate the identification and extraction of contract-related documentation requirements.

Objective 2: Predict and analyze the overhead costs and durations associated with contract-related documentation.

Objective 3: Develop novel approaches for improving and streamlining contract-related documentation and establishing performance indicators to quantify proposed improvements.

Objective 4: Quantify and streamline documentation workflows, with a focus on people and communication.

1.3 Scope of Research

In this thesis, the quantification and streamlining of documentation workflows by focusing on people and communication (Chapter 4) was limited to construction change order reports. Change orders, as one type of construction report, requires effective communication between diverse project participants to control delays and costs. Due to their significant impact on project cost and time, therefore, change orders were chosen as the document of study in the documentation workflow research.

1.4 Research Methodology

This research has developed an innovative framework for managing contract-related reporting documentation in construction. The framework is capable of facilitating quick access to reporting requirements and streamlining the processes associated with completing these requirements. The framework is designed to address the challenges in both the planning and the execution and control phases of contract-related reporting documentation processes. The framework employs the power of text-mining, machine learning, simulation, Lean techniques, and others to enhance contract-related reporting documentation in construction.

The proposed framework consists of four components designed to:

1. Identify required reporting documents using automated processes,
2. Estimate the time/cost of the required reporting documents,
3. Quantify and streamline the reporting documentation workflow by focusing on process design, and
4. Quantify and streamline the reporting documentation workflow by focusing on people and communication.

The proposed framework is illustrated in Figure 1.1.

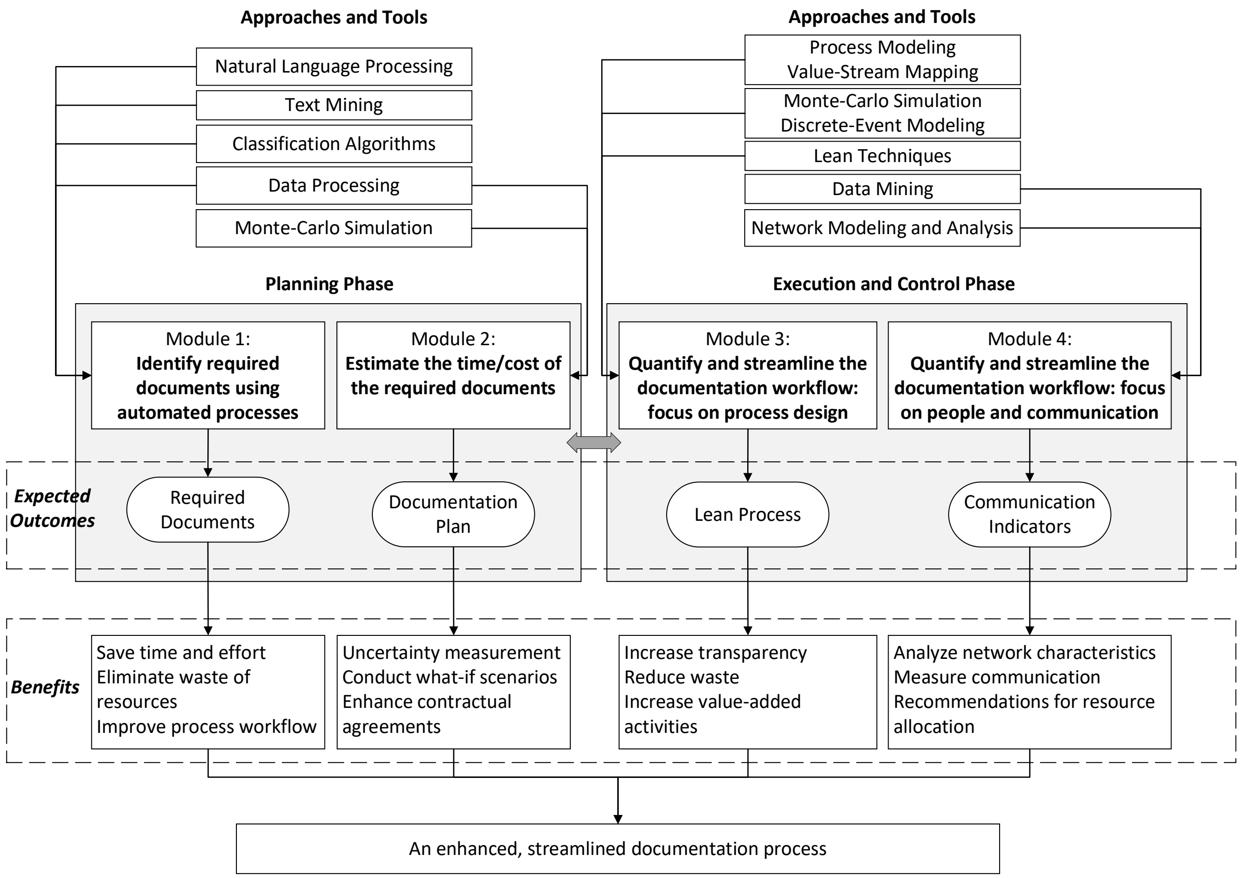


Figure 1.1: An overview of the proposed framework

In the planning phase, modules 1 and 2 focus on identifying the required contract-related reporting documentation and predicting associated effort (i.e., cost and durations) to facilitate the preparation of a documentation plan capable of addressing the requirements during the planning phase of the project. Here, a set of approaches and tools capable of automatically extracting reporting documentation requirements embedded in contract documents was used. Then, a Monte-Carlo simulation model was created and used to predict the overhead costs and durations associated with completing contract-related reporting documentation. Consideration of specific requirements during the preliminary planning stage can help ensure that (1) an adequate number of personnel are available to complete reporting documentation requirements on-time and within budget, (2) efficient project documentation systems are implemented, and (3) redundant and/or overlapping reporting documentation requirements are addressed prior to project execution. These modules are expected to assist decision-makers with the development of a documentation plan, including expected time and costs, efficient resource allocation, specialized data collection systems, and document template development to more efficiently fulfil client expectations and needs.

In the execution and control phase, module 3 is focused on establishing indicators for the control and implementation of construction documentation to measure the performance of the documentation process, to facilitate waste reduction in the construction administrative process, and to identify areas of improvement. Here, Lean techniques and simulation modeling were used to improve process design. Notably, module 3 can be iteratively repeated during the execution phase to promote a culture of continuous process improvement. Module 4 is focused on the participants of the reporting documentation process to assess their communication and quantify the documentation process networks. The communication of project participants is an important factor for efficient project documentation and should be assessed periodically during the execution of a project. Here, social network analysis was used to identify and provide recommendations to eliminate communication bottlenecks, thereby enhancing the documentation process. The use of multiple indicators to quantify process performance during the execution and control phase is expected to provide decision-makers with the information required to take necessary actions to streamline and smooth the workflow.

This framework provides practitioners and researchers with systematic guidelines to improve and enhance the documentation process to improve functionality and efficiency. Practical application of this approach is anticipated to provide decision-makers with the insights necessary to enhance contract negotiations, document workflow processes, submittal procedures between clients and contractors, and resource allocation, in turn increasing value for all project stakeholders. The feasibility and functionality of the proposed framework was validated using practical case studies, the results of which have also provided valuable information for practitioners.

1.5 Thesis Organization

This thesis is organized following a paper-based format that is consistent with the research framework shown in Figure 1.1. The chapters in this thesis are organized as follows:

Chapter 2 develops a novel methodology to automate the identification of contract-related documentation requirements and a method for predicting and analyzing the cost and durations associated with document preparation to enhance documentation planning. Documentation requirements are extracted using Natural Language Processing (NLP) and Machine Learning (ML), and stochastic simulations are then used to predict overhead costs and durations associated with document preparation.

Chapter 3 focuses on how Lean approaches and simulation modeling can be integrated to enhance the overall performance of the documentation process by focusing on process design. The proposed methodology is able to quantitatively assess the performance of the documentation process, to identify potential improvements to the current process, and to quantitatively predict the impact of proposed improvements on future project performance. Providing indicators for the control and implementation of construction documentation would increase transparency, to allow waste reduction in construction administrative processes, and to identify areas of improvement.

Chapter 4 focuses on the discovery and quantification of documentation process networks for improved efficiency. Social network analysis is employed to quantify and assess communication of project participants within documentation networks. Structural properties of various relationships in the construction documentation process were analyzed to ensure project members efficiently circulate documents in a timely manner. Communication networks of documentation process were discovered and modeled at various levels in the study. Then, structural characteristics of the communication network and various communication indicators were measured and analyzed to better understand the communication performance of project members and to identify communication bottlenecks in the process.

Chapter 5 summarizes the conclusions, research contributions, limitations, and envisioned future work of this thesis.

Chapter 2

Automated Extraction and Time-Cost Prediction of Contractual Reporting Requirements in Construction Using Natural Language Processing and Simulation

2.1 Introduction

Work within the construction industry is allocated through construction contracts (Shash and Habash, 2020), which include information such as instructions, definitions, supporting statements, and contractual requirements that detail the standards and project specifications of the client (Barlow *et al.*, 2014). A core component of construction contracts is reporting and information requirements, which require contractors to periodically submit various reports detailing different aspects of project progress to the client (El-Omari and Moselhi, 2011). As construction projects and contracts are becoming increasingly complex, clients are demanding that contractors provide more information and reports on different project aspects (Hassan and Le, 2020; Lee *et al.*, 2019). Reporting has quickly become a laborious procedure, with construction personnel spending as much as 40% of their time gathering field data, organizing and analyzing data, preparing reports, and verifying report accuracy (Jeong *et al.*, 2015).

Although a large administrative burden for contractors (Caldas *et al.*, 2002), the time and resources needed to complete reporting requirements—as well as the precise reporting requirements themselves—are often unknown and unaccounted for during the preliminary planning stages of a project. An integral component of project success, preliminary planning involves, among other activities, the selection of the project management team and the creation of the project documentation system. Consideration of specific reporting requirements during the preliminary planning stage ensures that (1) an adequate number of personnel is available to complete reporting requirements on time and within budget, (2) efficient project reporting systems are implemented, and (3) redundant and/or overlapping reporting requirements are addressed prior to the execution phase of a project. Contract documentation, however, remains an immature area of practice, with the identification of reporting requirements involving the manual reading, interpretation, and analysis of hundreds of unstructured textual contract pages to differentiate between statements related to requirements and other unimportant texts, such as instructions and definitions. Due to the time and effort involved, the manual extraction of reporting requirements is often not completed during the preliminary planning stages of a project, with project managers informally approximating reporting costs and resource requirements. Indeed, it has been reported that office-related processes, such as project reporting, continue to suffer from low reliability, where planned durations are often underestimated (Pestana *et al.*, 2014).

The poor estimation of project reporting costs and resource requirements during the preliminary planning stages of construction can result in a number of challenges for contractors (ElGindi, 2017; Hassan and Le, 2020; Jeon *et al.*, 2020; Levin, 2016). For example, an inadequate number of available project management personnel may result in project reports that are submitted late or with errors. In the case of certain types of contracts (e.g., cost-plus), reporting costs that exceed the preliminary estimate can result in disputes between the contractor and client. Furthermore, a lack of understanding of the reporting requirements in the preliminary stages of a project may prevent contractors from increasing reporting efficiency during the construction phase through the consolidation of redundant reporting requirements or by optimizing the composition of the project management team. As such, the ability to quickly, accurately, and automatically extract reporting requirements and predict associated costs is expected to have a notable impact on project performance (ElSawy *et al.*, 2011). Although text-mining techniques, such as information extraction, text classification, and other predictive analytics, have been used by researchers to develop requirement extraction models (Hassan and Le, 2020; Jallow *et al.*, 2017; Lee *et al.*, 2019), existing models are not fully automated, do not provide high requirement extraction accuracy, and lack a cost-and-time prediction component. Methods capable of automatically extracting contractual reporting requirements and predicting the time and costs associated with report preparation, therefore, remain relatively unexplored.

To address this challenge, this study has developed a framework capable of (1) automating the identification and extraction of reporting requirements and (2) predicting and analyzing the overhead costs and durations associated with report preparation. The framework employs Natural Language Processing (NLP) and Machine Learning (ML) techniques to automate the extraction of reporting requirements, and uses stochastic simulation to predict the durations and costs associated with report preparation using historical project data. Real contractual documents from an actual case study were used to (1) develop and refine the reporting requirement extraction module and (2) demonstrate the functionality and validity of the complete framework. This framework provides practitioners and researchers with an automated tool to more efficiently identify reporting requirements and quantify the time and costs associated with report preparation. Practical application of this approach is anticipated to provide decision-makers with the insights necessary to enhance contract negotiations, reporting workflow processes, and submittal procedures between clients and contractors, in turn increasing value for all project stakeholders.

2.2 Research Background

2.2.1 Construction Reporting

Many problems in the construction industry involve communication and reporting procedures, with ineffective reporting systems leading to poor project management (ElGindi, 2017; Jafari *et al.*, 2020; Morgan, 2010). Construction reports, therefore, are often required by clients as a means of monitoring project progress, estimating production rates, and resolving disputes and claims (Jeong *et al.*, 2015). Project reporting involves the collection and structuring of large volumes of site data from numerous field management activities by many site personnel on a frequent—even daily—basis (Lee *et al.*, 2020; Shrestha and Jeong, 2017). Given the amount of preparation work required together with the frequency of submittals, reporting has become a time- and effort-intensive procedure that can result in notable increases in overhead costs of the project.

Various construction field management tools have been developed to establish project reporting systems tailored to the needs of contractors, while ensuring the reporting requirements of projects are met (Barlow *et al.*, 2014; Jeong *et al.*, 2015; Lee *et al.*, 2020). For example, Russell (Russell, 1993) developed a daily construction project management system that rapidly reports and shares site information and project progress status between project participants. Similarly, Shiau and Wang (Shiau and Wang, 2003) developed a construction management information system consisting of daily reports as well as cost management and design-change management modules to compile daily site management information. El-Omari and Moselhi (El-Omari and Moselhi, 2011) proposed a model to facilitate automated data acquisition from construction sites by deploying an information technology platform. Their goal was to integrate automated data acquisition technologies to collect required data for progress measurement purposes to support efficient time–cost tracking and control of construction projects (El-Omari and Moselhi, 2011). Following the same line of work, Lee et al. (Lee *et al.*, 2020) proposed an approach to automatically generate daily reports from text messages exchanged through a commonly used text messaging system.

It is important to note that the aforementioned models were primarily focused on effective data acquisition, information flow, and communication to facilitate the monitoring of site work, incurred costs, and potential challenges (El-Omari and Moselhi, 2011; Omar and Nehdi, 2016). Although these studies have addressed the downstream aspect of reporting, they have been developed with the assumption that reporting requirements are already defined and known in advance. In practice, however, reporting requirements for complex types of construction, such as oil and gas or infrastructure projects, often differ between projects and from contract-to-contract, making the time, resources, and costs associated with report preparation difficult to approximate. Methods for automating the extraction of contractual reporting requirements or the estimation of time and cost implications associated with reporting, however, remain relatively unexplored.

Importantly, the lack of research literature in the area of contract documentation is not indicative of the practical importance of this issue. Discussions with experienced professionals at a construction company in Alberta, Canada, revealed that contractors are very interested in techniques that can support the extraction, management, and time–cost prediction of reporting requirements. Once considered an obligatory and static activity, contractors are beginning to explore methods capable of enhancing the planning, and therefore efficiency and cost, of project reporting—particularly for complex types of construction where contracts are often specific to each individual project.

2.2.2 NLP Applications in Construction

To avoid unnecessary changes, rework, and potential claims, contractors must thoroughly analyze construction contracts and specifications to ensure that client requirements are identified, managed, and fulfilled (Jallow *et al.*, 2017). The traditional approach of identifying reporting requirements involves the manual reading, interpretation, and analysis of hundreds of unstructured textual contract pages to differentiate between statements related to requirements and other extraneous text (e.g., instructions and definitions). Techniques capable of accelerating the reporting requirement extraction process, therefore, represent a key prerequisite for the development of an automated time–cost prediction model.

Natural Language Processing (NLP) is an area of Artificial Intelligence (AI) that focuses on the development of techniques to analyze, process, and extract information from natural human language. Applications include machine translation, speech recognition, and automated content analysis (Manning and Schutze, 1999). In construction, a large number of project documents are generated in text format (Tixier *et al.*, 2016). The use of NLP techniques to organize and improve access to information contained in these types of documents is becoming ever more essential for effective construction management (Caldas *et al.*, 2002), with NLP techniques being increasingly applied in construction research (Fan and Li, 2013; Tixier *et al.*, 2016; Zhang *et al.*, 2020). Caldas et al. (Caldas *et al.*, 2002), for instance, employed NLP techniques to automate the classification of construction documents to improve organization of and access to information within interorganizational systems. Al Qady and Kandil (Al Qady and Kandil, 2015) also developed an automated classification system of construction documents according to their semantic relationships. Fan and Li (Fan and Li, 2013) used NLP for the automatic retrieval of similar cases from an electronic case repository of construction accidents.

Text classification, a subfield of NLP, is an automated process for classifying text into categories (Manning *et al.*, 2008; Russell and Norvig, 2002). Text classification is divided into rule-based and Machine Learning (ML)-based methods (Manning *et al.*, 2008). Rule-based text classification categorizes text using a manually defined pattern to create rules; in contrast, under ML-based text classification, a machine learns how to classify text on its own using data. A variety of text classification models have been developed for the construction domain. For example, Salama and El-Gohary (Salama and El-Gohary, 2016) developed a hybrid semantic, multilabel ML-based text classification algorithm for classifying clauses and subclauses of general conditions to support automated compliance checking. Lee et al. (Lee *et al.*, 2019) proposed a rule-based model to automatically detect risk-related sentences of contracts to support contract risk management for construction contractors. Zhong et al. (Zhong *et al.*, 2020) combined NLP and convolutional neural networks to develop a classification model capable of automatically classifying accident narratives to support safety management on site. Zhou and El-Gohary (Zhou and El-Gohary, 2016) proposed an ontology-based, multilabel text classification approach for classifying environmental regulatory clauses to support automated compliance checking in construction, and Hassan and Le (Hassan and Le, 2020) proposed a domain-specific classification model to identify client requirements from construction contracts. It is important to note that the implementation of existing text classifiers to different applications remains limited, as text classification models, text features, and performance requirements vary greatly across domains and applications (Salama and El-Gohary, 2016). Designed for a specific domain, the aforementioned text classifiers and are not well-suited for applications that require alternate classification structures.

2.2.3 Research Gap

Despite these advancements, research focused on enhancing the management of contractual reporting requirements remains relatively unexplored and fragmented. Most of the studies in the area of construction reporting have focused on the development of systems that improve data acquisition, information flow, and communication. While other studies, such as those mentioned previously, have deployed NLP and AI to automate information retrieval and extraction from construction contracts, the text approaches used to develop these requirement extraction models are limited by a lack of full automation, low extraction accuracy, and the absence of a cost–time prediction component (Hassan and Le, 2020; Jallow *et al.*, 2017; Lee *et al.*, 2019). Indeed, a review of construction literature could not identify any established study capable of automatically extracting reporting requirement statements from hundreds of pages of contractual and project specification documents.

2.3 Framework Overview

To address the gap existing in literature, this study developed a novel, NLP-based framework for the automated extraction and time–cost prediction of contractual reporting requirements in construction. The framework consists of two modules, namely the (1) automatic extraction of reporting requirements module and (2) prediction of reporting time and cost module that are linked as illustrated in Figure 2.1.

The first module, hereafter referred to as the extraction module, is responsible for identifying statements in contract or project specification documents that prescribe reporting requirements. Contract documents and project specifications are input into the NLP model. Relevant text is extracted from the documents and is transformed into a format that is compatible with the text classification models.

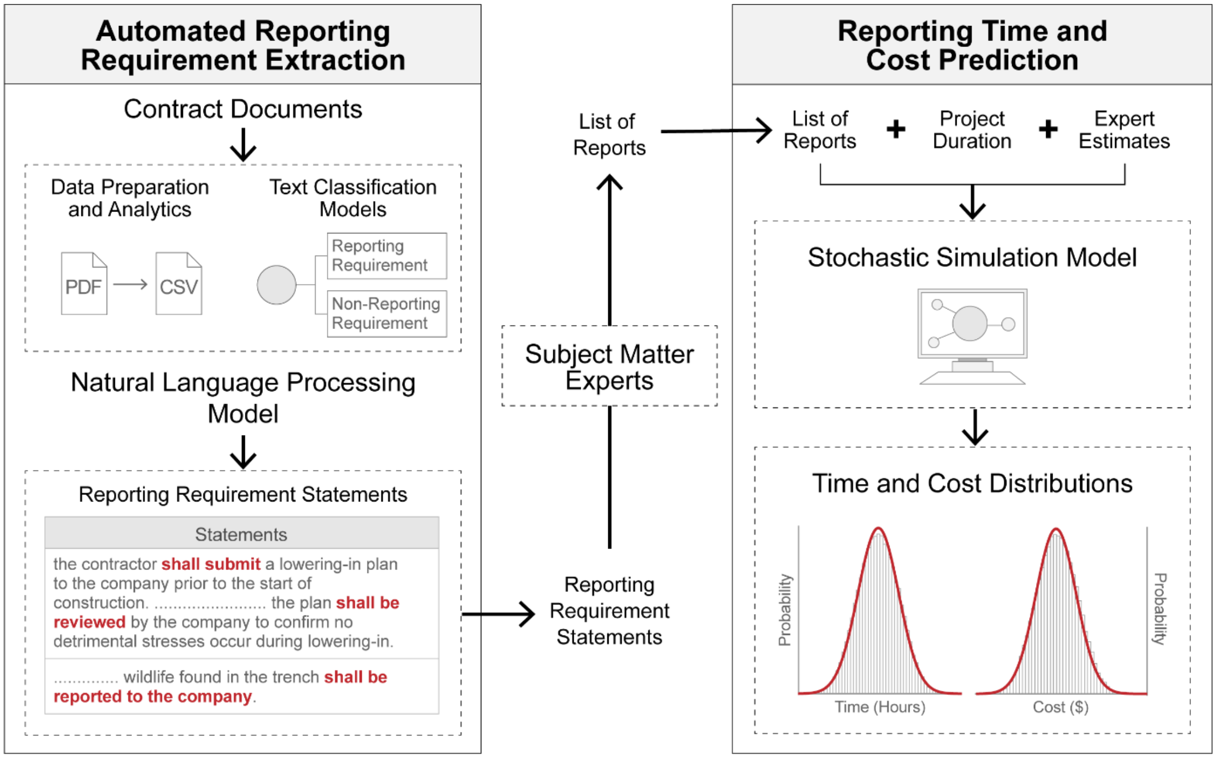


Figure 2.1: Proposed framework for automated requirement extraction and time/cost prediction

Text classification algorithms are then used to classify contractual and project specification statements as either a (1) reporting requirement or (2) non-reporting statement. Both rule-based and ML-based text classification methods can be used to classify statements; application of either method will depend on the specific requirements of the user. While rule-based text classification is more time-consuming than ML-based classification due to the involvement of manual rule development, rule-based classification commonly results in higher precision and recall (Manning *et al.*, 2008). ML, on the other hand, makes it possible to automatically classify text, provided that sufficient learning opportunities are available (Russell and Norvig, 2002).

The second module, hereafter referred to as the prediction module, is responsible for generating relevant time–cost predictions. Reporting requirements output from the extraction module are used by practitioners to prepare a list of required reports and their associated submittal frequencies that are then input into the prediction module. Estimates of the time required to prepare a specific report are used as inputs. Then, a Monte Carlo simulation model, which uses random sampling to obtain numerical results or a probability distribution (Hastings, 1970), is used to predict the cost associated with report preparation based on project duration and historical data.

Although distinct, the practical functionality of these two modules increases considerably when used together. Manual extraction is very tedious and time-consuming, and contractors do not have enough time during the bidding stage to identify the contract requirements and plan accordingly. Because of the ability of the extraction module to quickly extract reporting requirements, the prediction module can now be applied in a more impactful stage of the project life-cycle (i.e., pre-construction bidding and planning stages). Specifically, the outputs of the prediction module can be used to (1) ensure that sufficient cost and time contingencies for report preparation are included in bid estimates, (2) engage in negotiations to reduce redundant reporting requirements before finalizing contracts, and (3) improve resource allocation.

2.4 Extraction Module

Development of the extraction module was completed in three main steps, namely (1) data preparation and pre-processing, (2) development and training of text classification models, and (3) evaluation of model performance. Ten contractual and specification documents of an oil-and-gas project were supplied by a private Canadian construction contracting firm and were used to develop the extraction module. Python (Python, 2018), an open-source programming language, was used to automate module development steps.

2.4.1 Data Preparation and Pre-processing

Data preparation and pre-processing transformed raw data into a labeled dataset that was used to develop and train the text classification models. This involved the (1) extraction of textual data from documents, (2) manual assignment of labels to extracted statements, and (3) cleaning of labeled data. An overview of the data preparation process is illustrated in Figure 2.2.

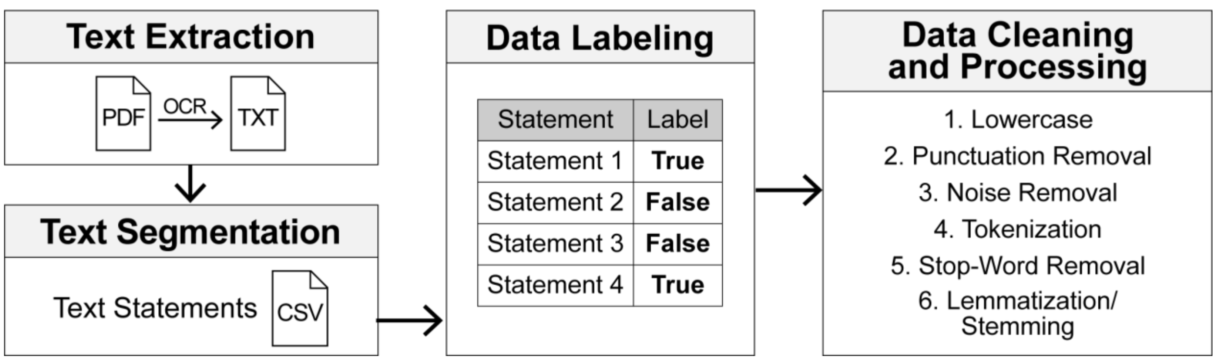
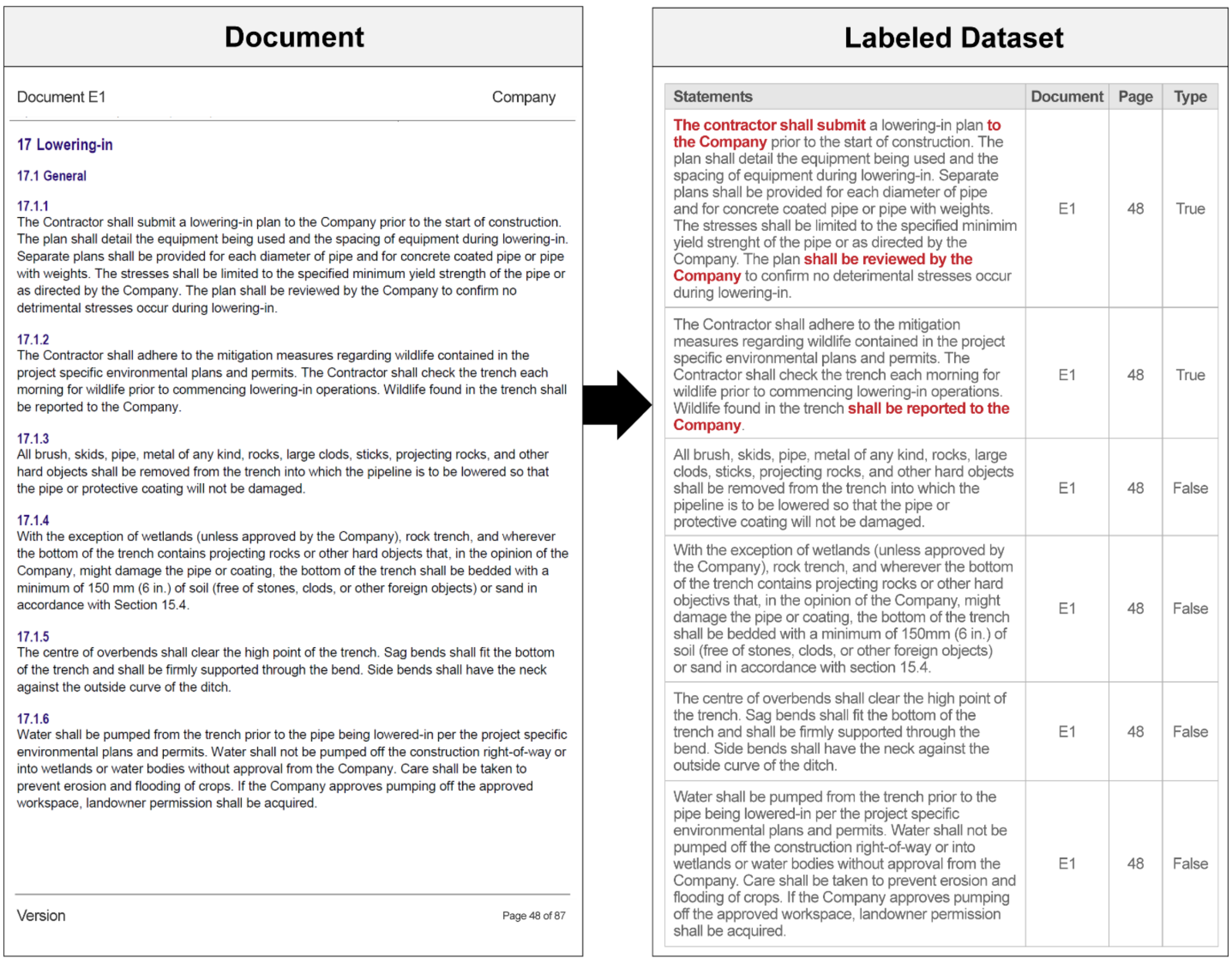


Figure 2.2: Data preparation and pre-processing

Documents, which were provided in an imaged portable document format (i.e., .pdf), were first converted into a standard, processable text format (i.e., .txt) using Optical Character Recognition (OCR). Next, text documents were automatically segmented into individual text statements using document formatting. A total of 8943 text segments were extracted from 10 contractual and project specification documents. Since pre-labeled textual construction datasets are not as readily available as other domain applications, such as movie reviews or twitter messages (Priyanka *et al.*, 2019), extracted statements were labeled manually. A label of "true" was manually applied to statements prescribing a reporting requirement, while a label of "false" was applied to non-reporting statements. A sample of the labeled dataset is presented in Figure 2.3.



Document

Document E1 Company

17 Lowering-in

17.1 General

17.1.1
The Contractor shall submit a lowering-in plan to the Company prior to the start of construction. The plan shall detail the equipment being used and the spacing of equipment during lowering-in. Separate plans shall be provided for each diameter of pipe and for concrete coated pipe or pipe with weights. The stresses shall be limited to the specified minimum yield strength of the pipe or as directed by the Company. The plan shall be reviewed by the Company to confirm no detrimental stresses occur during lowering-in.

17.1.2
The Contractor shall adhere to the mitigation measures regarding wildlife contained in the project specific environmental plans and permits. The Contractor shall check the trench each morning for wildlife prior to commencing lowering-in operations. Wildlife found in the trench shall be reported to the Company.

17.1.3
All brush, skids, pipe, metal of any kind, rocks, large clods, sticks, projecting rocks, and other hard objects shall be removed from the trench into which the pipeline is to be lowered so that the pipe or protective coating will not be damaged.

17.1.4
With the exception of wetlands (unless approved by the Company), rock trench, and wherever the bottom of the trench contains projecting rocks or other hard objects that, in the opinion of the Company, might damage the pipe or coating, the bottom of the trench shall be bedded with a minimum of 150 mm (6 in.) of soil (free of stones, clods, or other foreign objects) or sand in accordance with Section 15.4.

17.1.5
The centre of overbends shall clear the high point of the trench. Sag bends shall fit the bottom of the trench and shall be firmly supported through the bend. Side bends shall have the neck against the outside curve of the ditch.

17.1.6
Water shall be pumped from the trench prior to the pipe being lowered-in per the project specific environmental plans and permits. Water shall not be pumped off the construction right-of-way or into wetlands or water bodies without approval from the Company. Care shall be taken to prevent erosion and flooding of crops. If the Company approves pumping off the approved workspace, landowner permission shall be acquired.

Version Page 48 of 87

Labeled Dataset

Statements	Document	Page	Type
The contractor shall submit a lowering-in plan **to the Company** prior to the start of construction. The plan shall detail the equipment being used and the spacing of equipment during lowering-in. Separate plans shall be provided for each diameter of pipe and for concrete coated pipe or pipe with weights. The stresses shall be limited to the specified minimim yield strenght of the pipe or as directed by the Company. The plan **shall be reviewed by the Company** to confirm no deteremental stresses occur during lowering-in.	E1	48	True
The Contractor shall adhere to the mitigation measures regarding wildlife contained in the project specific environmental plans and permits. The Contractor shall check the trench each morning for wildlife prior to commencing lowering-in operations. Wildlife found in the trench **shall be reported to the Company**.	E1	48	True
All brush, skids, pipe, metal of any kind, rocks, large clods, sticks, projecting rocks, and other hard objects shall be removed from the trench into which the pipeline is to be lowered so that the pipe or protective coating will not be damaged.	E1	48	False
With the exception of wetlands (unless approved by the Company), rock trench, and wherever the bottom of the trench contains projecting rocks or other hard objectivs that, in the opinion of the Company, might damage the pipe or coating, the bottom of the trench shall be bedded with a minimum of 150mm (6 in.) of soil (free of stones, clods, or other foreign objects) or sand in accordance with section 15.4.	E1	48	False
The centre of overbends shall clear the high point of the trench. Sag bends shall fit the bottom of the trench and shall be firmly supported through the bend. Side bends shall have the neck against the outside curve of the ditch.	E1	48	False
Water shall be pumped from the trench prior to the pipe being lowered-in per the project specific environmental plans and permits. Water shall not be pumped off the construction right-of-way or into wetlands or water bodies without approval from the Company. Care shall be taken to prevent erosion and flooding of crops. If the Company approves pumping off the approved workspace, landowner permission shall be acquired.	E1	48	False

Figure 2.3: Sample contract document (*left*) and annotated dataset (*right*)

Data were then structured into a single, comma-separated values (.csv) file, with text statements stored in the first column and the document name and associated page stored in subsequent columns. The last column contained the pair label (i.e., type: true/false) of each statement. The labeled dataset was validated by domain experts to ensure accuracy of the manual labels. The final dataset used in the study included 340 reporting requirements and 8603 non-reporting statements.

The final dataset was then cleaned to reduce data noise and enhance the quality of data used to train the text classification models. First, text was converted to a lowercase form to ensure identical words were treated as like terms (e.g., "Submit" and "SUBMIT"). Then, punctuation was removed from the text. In addition to text in the main body of the document, OCR extracted text from footers, page numbers, headers, annotations, and footnotes. This text acted as data noise for the text classification algorithms and was, therefore, removed. Then, tokenization was used to divide text statements into words (i.e., tokens) and to convert text into a feature vector form to prepare text for feature engineering and further analysis (Grefenstette and Tapanainen, 1994). Stop-words, which are frequent words such as conjunctions, prepositions, and pronouns (e.g., the, for, so, is, of, and a) that do not carry relevant information for text classification, were removed using a standard English stop-word list (Manning and Schutze, 1999). Lemmatization and stemming were applied to reduce the number of features through word grouping. While word stemming groups words by removing prefixes and/or suffixes to conflate words to their original root (Porter, 1980), lemmatization groups words subsequent to a full morphological analysis. Once data cleaning was completed, the dataset was randomly split into training (80%) and testing (20%) sets, which were used to train the text classifiers and to evaluate classifier performance, respectively.

2.4.2 Rule-Based Classification

In the rule-based classification approach, a set of hand-coded "IF-THEN" rules that define the label assignment criteria for a certain category were prepared (Lee *et al.*, 2019). These rules were iteratively constructed and refined to improve accuracy of the classifier. The process used to develop the rule-based classification model is depicted in Figure 2.4.

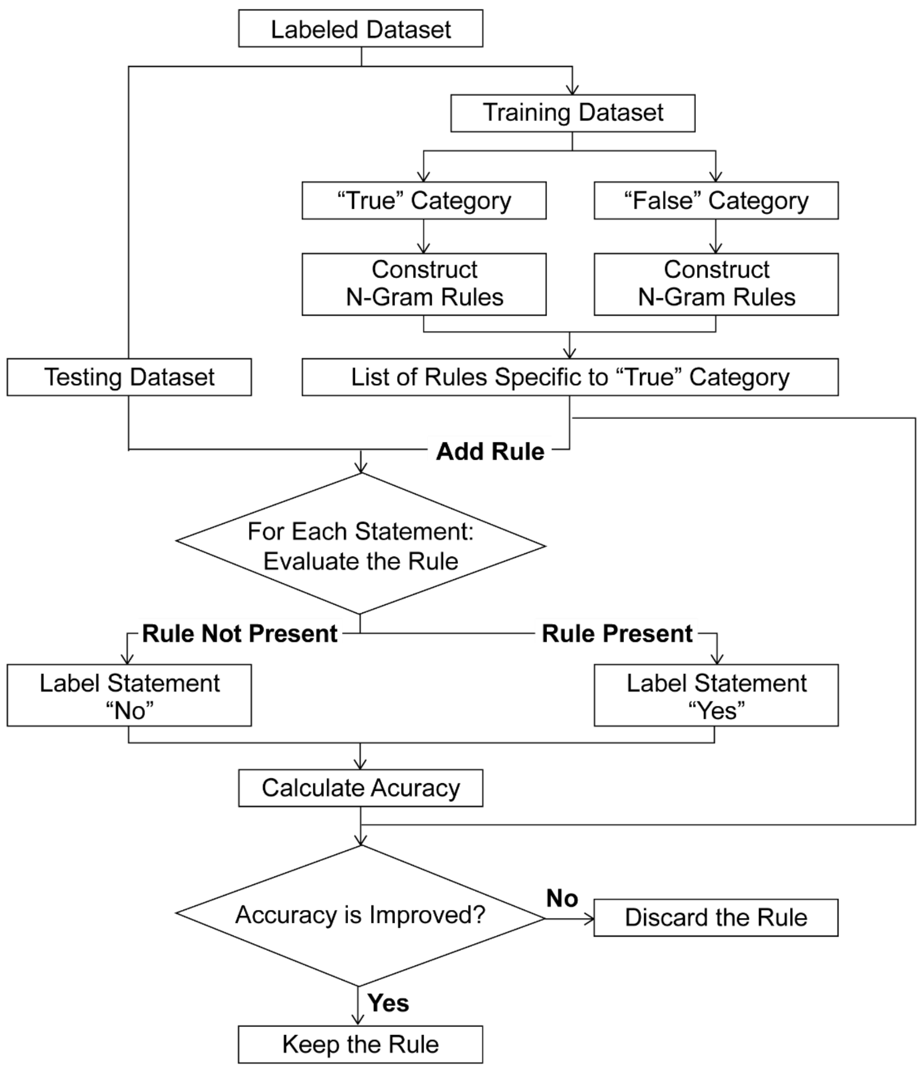


Figure 2.4: Rule-based classification

Accurate filtering of reporting requirements from contractual documents requires the development of robust and comprehensive rules. Although keywords, such as “report” and “submit,” may be helpful in identifying certain reporting requirements, construction contracts also contain key phrases, such as “shall be reported/submitted,” which indicate that a report or deliverable must be provided contractually. It is important to note that keywords alone cannot distinguish a reporting requirement from any other contractual requirement. As such, critical phrases were extracted using text analytics. Using the training dataset, the rule-based model was used to extract n-grams (i.e., a sequence of co-occurring words as a single token) from the textual statements. The most common n-grams (i.e., phrases) appearing in the reporting requirement statements are summarized in Figure 2.5.

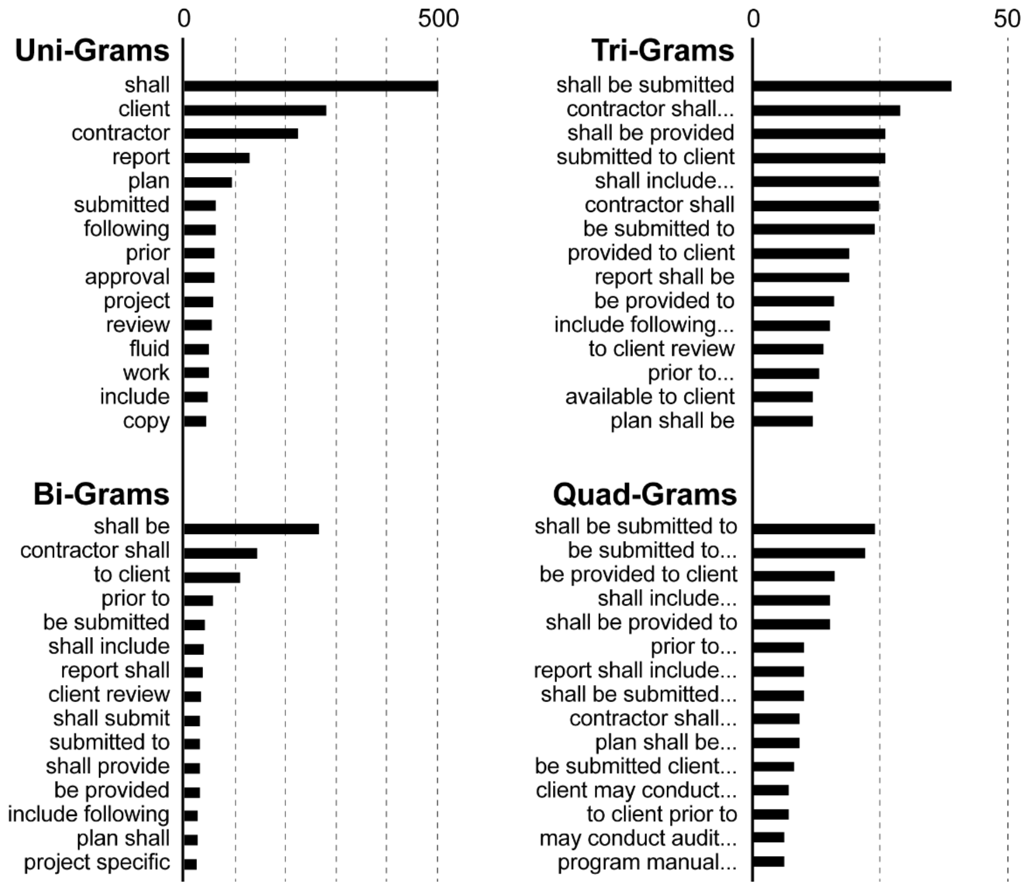


Figure 2.5: Most common n-grams in reporting requirement statements

To avoid errors, a list of n-grams specific only to the “true” category was prepared. N-grams common to both the “true” and “false” categories were removed. The final rules consisted of four different sets of n-grams capable of discerning between “true” and “false” statements, namely uni-grams, bi-grams, tri-grams, and quad-grams representing single-, two-, three-, or four-word phrases, respectively. These four sets of n-grams were developed to evaluate the effect of each n-gram set on the performance of the rule-based text classifier. N-grams were flagged as rules, with each rule consisting of a pattern and a predicted category. N-grams in each N-gram set were closely monitored, and rules for each statement were evaluated. Each rule was added, one-by-one, to the text classification model. Predicted labels were then compared to actual labels, and classifier performance was calculated. If the performance (i.e., accuracy) increased with the addition of the rule, the rule was retained. If not, the rule was removed. This was repeated for each rule of each n-gram until a final list was created.

2.4.2.1 Performance of Rule-Based Classification Models

Performance of the text classification models were evaluated using accuracy, precision, recall, and F1-score, which were calculated using Equation (2.1) through Equation (2.4), respectively,

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{2.1}$$

$$Recall = \frac{TP}{TP+FN} \tag{2.2}$$

$$Precision = \frac{TP}{TP+FP} \tag{2.3}$$

$$F1\text{- }score = \frac{2\times Precision\times recall}{Precision+recall} \tag{2.4}$$

where *TP* are true positives (i.e., statements correctly labeled ‘true’), *FP* are false positives (i.e., statements incorrectly labeled ‘true’), *TN* are true negatives (i.e., statements correctly labeled ‘false’), and *FN* are false negatives (i.e., statements incorrectly labeled ‘false’).

Accuracy (Equation (2.1)) is defined as the percentage of correctly classified statements over the total number of statements in the testing set. Recall (Equation (2.2)), is defined as the percentage of true positives identified by the model. Precision (Equation (2.3)) is defined as the percentage of positives that are correctly labeled (Buckland and Gey, 1994). Finally, the F1-score (Equation (2.4)), combines precision and recall to provide an overall assessment of model effectiveness. Recall is considered to be the most critical performance metric in the context of requirement extraction, where the extraction of all reporting requirements is the primary objective. For instance, a model may have low performance accuracy because it results in a larger number of false positives (i.e., non-reporting statements labeled as requirements). However, the model may have high recall results (i.e., 100%) if it is able to correctly label all reporting requirements as 'true.'

Specific rules for text processing were developed and applied to improve results of the rule-based classification model. Initial tests were conducted on different n-gram sets. The testing approach was conducted in an iterative manner, and results from 24 different combinations of n-grams and text pre-processing techniques (e.g., stop-word removal, lemmatization, etc.) were compared. The four sets of n-grams extracted from the training set are summarized in Table 2.1. The total number of rules generated increased with the number of n-grams (Table 2.1). Using the process summarized in Figure 2.4, the number of rules maintained for each n-gram was considerably reduced for all n-gram sets (Table 2.1). For example, of the 2363 rules generated for the bi-grams set, only 38 rules were retained. Accuracy was increased from 97%, using uni-grams, to 99%, using bi-grams, yet was decreased to 98% and 97% using tri- and quad-grams, respectively. Although differences between n-gram sets were minimal, optimal accuracy was achieved using the retained bi-gram rules. The impact of adding the first 10 and the last bi-gram rule on model accuracy is visualized in Figure 2.6.

Table 2.1: Number of generated and retained bi-gram rules and associated accuracy

N-Gram Set	Number of Generated Rules	Number of Retained Rules	Maximum Accuracy (%)
Uni-Grams	118	8	97
Bi-Grams	2,363	38	99
Tri-Grams	3,762	38	98
Quad-Grams	4,268	34	97

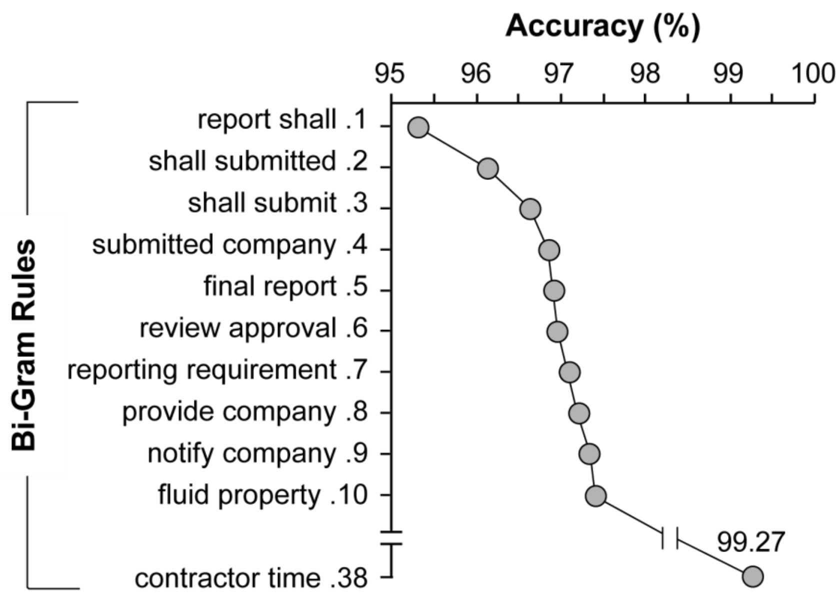


Figure 2.6: Impact of adding bi-gram rules on the accuracy of the rule-based text classifier

The first bi-gram rule, “report shall”, resulted in an accuracy of 95.3%. The third bi-gram, “shall submit”, further increased the accuracy of the classifier to 96.7%. The 37 rules added after “shall submit” collectively increased performance by 3.97% to 99.3%.

The impact of stop-word removal, lemmatization, and stemming on the performance of text classification models is known to differ based on the textual context and application. As such, the impact of stop-word removal and lemmatization/stemming were evaluated. Various experimental scenarios examining the impact of n-gram sets, stop-word removal, and lemmatization on model performance are summarized in Table 2.2.

Uni-grams had the lowest performance for both experimental scenarios (Table 2.2), and bi-grams demonstrated the highest performance in all three metrics in the base condition (Scenario 1).

Table 2.2: Effect of n-gram sets and data pre-processing on performance of rule-based classification for two experimental scenarios

N-Gram Set	**Class Label**	**Scenario 1:** *Without Lemmatization Stop-Words Retained*			**Scenario 2:** *With Lemmatization Stop-Words Removed*		
		Precision (%)	**Recall (%)**	**F1-Score (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Uni	True	91	56	69	93	55	69
	False	98	100	99	98	100	99
Bi	True	100	86	92	96	88	92
	False	99	100	100	99	100	100
Tri	True	99	86	92	98	71	83
	False	99	100	100	98	100	99
Quad	True	100	74	85	100	49	49
	False	99	100	99	97	100	99

When stop-words were removed and lemmatization was applied, bi-grams had the highest recall and F1-score, with precision only differing marginally from other n-grams. Interestingly, lemmatization and stop-word removal resulted in a 2% increase in the recall of the bi-gram classifier, while the recall of the other n-gram sets decreased (Table 2.2). Notably, the F1-score of tri-grams (without lemmatization and with stop-words retained) was equal to the F1-score of bi-grams (with lemmatization and with stop-words removed). This result is expected as, in some cases, removing stop-words from tri-grams transforms them into bi-grams. For example, when the stop-word "be" is removed from the tri-gram "shall be submitted", the bi-gram "shall submitted" results. Given the importance of the recall measurement when extracting reporting requirements, and based on the findings that bi-grams resulted in the highest model accuracy (Table 2.1) and recall (Table 2.2), bi-grams are selected as the optimal classifier for rule-based text classification.

2.4.3 Machine Learning-Based Classification

In contrast to rule-based classification, the alternate classification approach used in the present study was supervised ML models, with the learning process driven by previous knowledge of the data (Salama and El-Gohary, 2016). In ML algorithms, a general inductive process automatically builds a classification model for each class by observing the characteristics of a set of manually classified statements. The ML-based text classification approach is summarized in Figure 2.7.

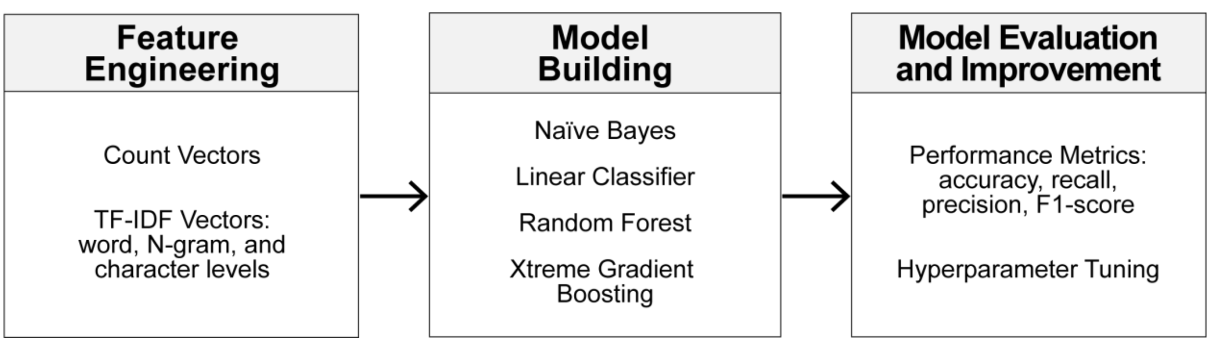


Figure 2.7: Machine learning-based text classification method

To ensure compatibility with computer processors, words were first converted into a numeric format using feature engineering. Here, raw text data were transformed into feature vectors, and new features were created using the dataset. Different methods were used to create relevant dataset features prior to input into the text classification algorithm (Forman, 2003).

Two approaches for constructing representation vectors, namely count vectors and term frequency-inverse document frequency (TF-IDF) vectors, were implemented. Count vectors are a matrix representation of the dataset, where every row represents a statement, every column represents a word, and every cell represents the frequency count of a particular word (i.e., either zero or a real number) in a particular statement (Sebastiani, 2002). Words that appear in many textual statements are considered less meaningful and, therefore, each vector component (i.e., a word) can be weighed based on the number of statements in which the word appears. Another approach for constructing representation vectors is TF-IDF, which is a technique designed to identify important terms in a dataset by weighing a term's frequency (TF) together with its inverse document frequency (IDF), which weighs down high-frequency domain-specific terms while scaling up rare terms (Sebastiani, 2002). In TF-IDF vectors, terms can be extended to include characters and n-gram-level models, such as uni-gram (i.e., words), bi-grams (i.e., pairs of words), as well as tri and quad-grams (i.e., phrases). The TF-IDF of terms are calculated using Equation (2.5) (Sebastiani, 2002),

$$TF\text{-}IDF = \frac{tf_i}{T} \times \left(1 + log\left(\frac{N}{N_i}\right)\right) \tag{2.5}$$

where tf_i is the frequency of term i in the statement, T is the total number of words in the statement, N is the total number of statements, and N_i is the number of statements containing term i.

For the proposed method to be feasible in practice, retraining and prediction (Valieva *et al.*, 2021) must be completed within a relatively short period of time. As such, models that required longer than an hour to be fine-tuned and retrained (e.g., deep learning algorithms) were excluded from this study to ensure applicability of this research. Based on this criterion, four popular supervised ML algorithms, which have been shown to perform differently depending on the application and domain (Hassan and Le, 2020; Salama and El-Gohary, 2016), were implemented to build the ML-based text classification model. Characteristics of the ML algorithms are summarized as follows:

Naïve Bayes (NB) is a simple algorithm, based on the Naïve Bayes Theorem, that is used for solving practical domain problems including text classification (Witten *et al.*, 2016). Because it assumes that every feature is conditionally independent of other features for a given class label, computational cost of applying the NB algorithm is comparatively low.

Logistic Regression (LR) is a linear statistical ML algorithm that correlates discrete categorical dependent features with a set of target variables (Witten *et al.*, 2016). It is a complex form of linear regression that can predict data probability for predefined categories.

Random Forest (RF) is a supervised ML method based on ensemble learning that involves the construction of multiple decision trees during training. Outputs are classes that are averaged or voted the most by individual trees (Breiman, 2001). Decision Tree algorithms, such as the RF classifier, are often used to combat imbalanced classes, such as the scenario described here, where the number of non-reporting statements considerably exceeds the number of reporting requirements.

Extreme Gradient Boosting (XGBOOST) is a scalable ML system based on gradient boosting

(Chen and Guestrin, 2016). It generates a strong classifier by iteratively updating parameters of the former classifier to decrease the gradient of loss function. XGBOOST has superior performance in supervised ML, with high accuracy and low risk of overfitting.

2.4.3.1 Performance of ML-Based Classification Models

The final step in the development of the extraction module was the evaluation of the various ML-based text classification models. The performance of ML-based text classification algorithms is highly dependent on feature selection (i.e., domain dependent), type of ML techniques, and training datasets (Salama and El-Gohary, 2016). Therefore, all possible combinations of text pre-processing, feature engineering, and ML algorithms—resulting in 160 exhaustive combinations—were evaluated. Various conditions of text pre-processing approaches, such as stop-word retention or removal with or without the implementation of stemming and/or lemmatization, were tested to evaluate the effect on model performance. While the methodology was conducted in an iterative manner to allow for the detailed comparison of results, only a subset of the results is presented to maintain brevity.

The effect of using lemmatization or stemming is illustrated in Figure 2.8. Stemming improved classification accuracy of LR and XGBOOST algorithms, while lemmatization marginally improved the accuracy of NB and RF algorithms. Notably, the difference in classification performance accuracy between the two text pre-processing techniques was negligible, ranging from 0.03% to 0.4% (Figure 2.8). The XGBOOST algorithm with stemming applied resulted in the highest accuracy (98.4%).

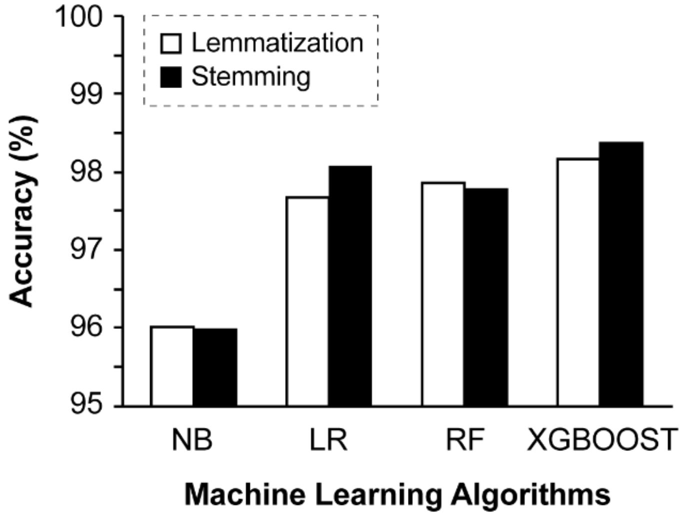


Figure 2.8: Impact of lemmatization and stemming on performance of Naïve Bayes (NB), Logistic Regression (LR), Random Forest, and Extreme Gradient Boosting (XGBOOST)-based machine learning algorithms

The recall, precision, and F1-score of the different ML algorithms were evaluated (Figure 2.9). Given that XGBOOST was found to have the highest accuracy with stemming, stemming was applied to all ML techniques prior to performance metric evaluation. The ML algorithms exhibited relatively similar recall values of over 98% for non-reporting statements (i.e., “false”).

In contrast, recall values for reporting requirements (i.e., “true”) varied considerably amongst the various ML algorithms. The NB algorithm resulted in the lowest “true” recall value (66%), whereas LR, RF, and XGBOOST algorithms resulted in “true” recall values of 74%, 74%, and 81%, respectively. The lower recall values and increased variability observed for the “true” class is likely due to the imbalanced distribution of statements in the contractual documents used (340 “true” requirements versus 8603 “false” statements). In terms of precision, RF, LR, and XGBOOST resulted in “true” precision results of 100%, 87%, and 87%, respectively. The

XGBOOST algorithm exhibited the highest recall (Figure 2.9) and accuracy (Figure 2.8) results for both the “true” and “false” classes and the second-highest F1-score and precision measurements.

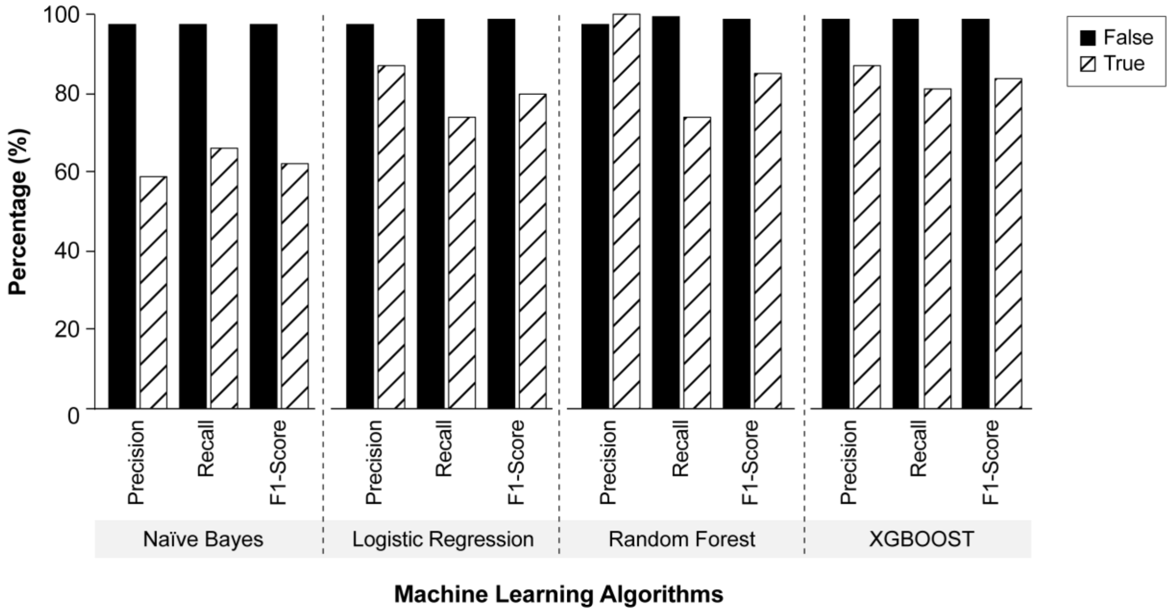


Figure 2.9: Performance measures of machine learning algorithms using uni-gram text classifications and word stemming

Variations in recall when using different n-gram sets for both the “true” and “false” class were evaluated and illustrated in Figure 2.10. As discussed previously, higher recall values were observed for non-reporting statements (i.e., “false” class). Uni-grams resulted in higher “true” recall values compared to bi-grams for all classification algorithms except the NB algorithm. The combined use of uni-grams and bi-grams with the LR and XGBOOST classification models yielded the highest “true” recall performance, with values of 77% and 87%, respectively.

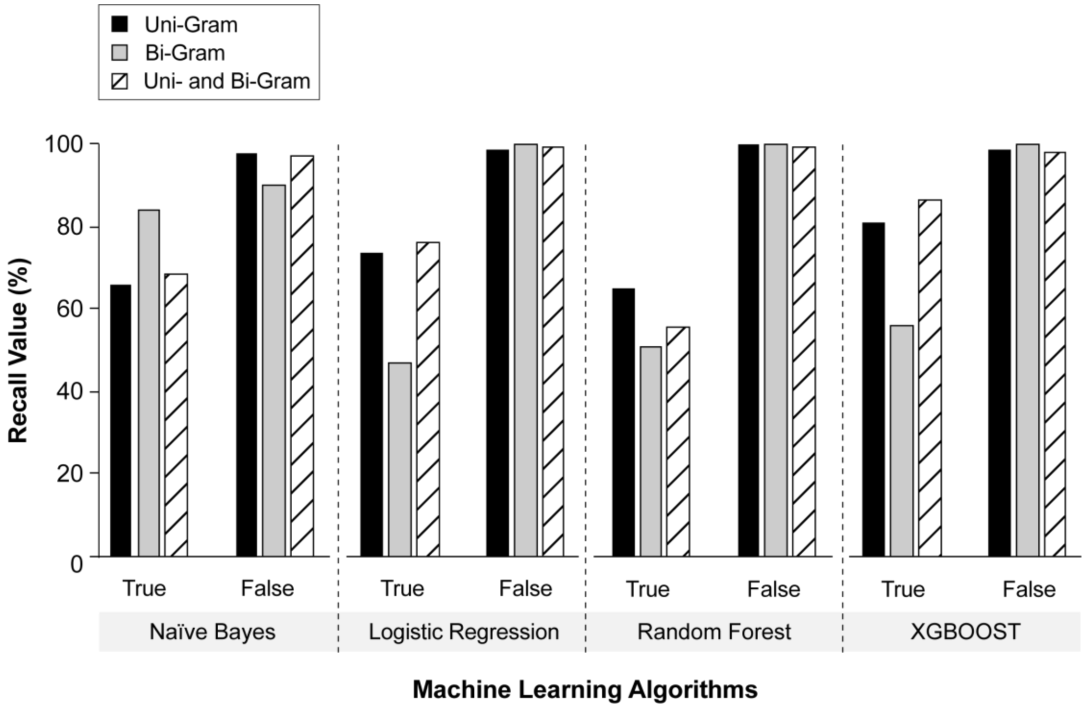


Figure 2.10: Impact of n-gram sets on recall performance of machine learning models

It is important to note, however, that a number of factors, such as dataset size, can influence the performance of ML models. The hyper-parameters of the ML models, therefore, must be tuned to specific data (Bergstra and Bengio, 2012). Here, hyperparameters were objectively changed, one-by-one, to mitigate overfitting and improve classifier performance. After identifying optimal hyperparameters (i.e., a single set of well-performing hyperparameters), the model was retrained with the full training dataset, and the testing dataset was re-evaluated.

The two models that were examined were the RF and XGBOOST algorithms, as they have many parameters, and the impact of their hyperparameters is significant. Table 2.3 summarizes the values of the four performance metrics of these two classifiers before and after fine-tuning of their hyperparameters. Fine-tuning parameters improved recall, precision, and F1-score measurements for both classifiers under both classes. The largest improvement for both the RF and XGBOOST models was observed for the "true" class. XGBOOST achieved the highest recall and F1-scores after fine-tuning for both the "true" and "false" classes at 89% and 100% for recall and 92% and 99% for F1-score, respectively. The results demonstrated that fine-tuning hyper-parameters to optimize parameter value by analyzing their impact, in terms of over- and underfitting, results in increased model robustness.

Table 2.3: Effect of fine-tuning hyper-parameters on performance metrics of Random Forest (RF) and Extreme Gradient Boosting (XGBOOST)-based machine learning algorithms

Metrics	**Class Label**	**RF**		**XGBOOST**	
		Before	**After**	**Before**	**After**
Accuracy	-	97.8	98	98.4	98
Precision	True	99	100	87	96
	False	98	98	99	99
Recall	True	59	74	81	89
	False	100	100	99	100
F1-score	True	73	85	84	92
	False	99	99	99	99

Altogether, ML-based performance measurements revealed that the XGBOOST model outperformed the other ML algorithms in terms of accuracy (Figure 2.8) and recall (Figure 2.10) performance. Accordingly, the XGBOOST model is selected as the optimal classifier for ML-based text classification.

2.4.4 Comparison of Classification Models

Performance results achieved by the best-performing rule-based and ML-based classifiers are summarized in Table 2.4. Under the rule-based classifier, application of the bi-gram rule set with lemmatization and stop-word removal resulted in accuracy and "true" recall values of 99% and 88%, respectively (Table 2.4). Comparatively, application of the XGBOOST-based machine learning algorithm resulted in accuracy and "true" recall values of 98% and 89%, respectively.

Table 2.4: Performance of rule-based versus machine learning-based text classification models

Metrics	**Class Label**	**Rule-Based**	**ML-Based**
Accuracy	-	99	98
Precision	True	96	96
	False	99	99
Recall	True	88	89
	False	100	100
F1-Score	True	92	92
	False	100	99

The patterns used to construct the rules in the rule-based model were manually defined, requiring more effort in terms of rule preparation. It is also important to note that the results of the rule-based model are quite sensitive to input rules: adding or removing a specific rule may have a considerable impact on classifier performance. Alternatively, the ML-based text classification model learns the classification process by using training data. In this regard, the results of the ML model are sensitive to the number of training sets, performing best in the presence of more data. Given the results provided in Table 2.4, the choice of classification model depends on the availability of training data for the ML- based model or the level of effort able to be invested for rule construction in the rule-based model.

2.5 Prediction Module

The prediction module is used to estimate the time, resources, and cost needed to fulfill the reporting requirements. Module inputs include (1) the list of reports prepared by subject-matter experts using outputs of the rule-based or ML-based extraction module that describe the types and submittal frequencies of the reporting requirements, (2) the resources, time, and hourly rate associated with each reporting requirement, and (3) estimated project duration. To account for underlying uncertainties in model inputs and outputs, a Monte Carlo simulation model is employed, with uncertain parameters (e.g., report preparation time and project duration) input as probabilistic distributions derived from historical data. If sufficient historical data are unavailable, probabilistic distributions, such as a triangular distribution with minimum, most likely, and maximum values reported by experts, can be input into the model instead (AbouRizk and Halpin, 1992). The Monte Carlo simulation is then run for multiple iterations, with each iteration randomly selecting a value from each parameter's distribution. Outputs of the model include the predicted (1) time, (2) cost, and (3) distribution among the various personnel types to complete the reporting requirements.

2.6 Case Study

An oil-and-gas project led by a private Canadian construction contracting firm was used to demonstrate the proposed framework. The project was considered a small-size project by the contractor and was awarded by the client to the contractor through a cost-plus contract type. This project was completed before the initiation of this research study. Actual durations of report preparation were not recorded by the contracting firm.

2.6.1 Data Collection

While manual extraction of reporting requirements is not required for future construction contracts, manual extraction was required, here, for initial development of the extraction module. As such, and for this case study only, the list of manually extracted reporting requirements from the set of contract documents detailed in Section 2.4.1 were input into the model. Notably, outputs of the rule-based or ML-based extraction module can be used to prepare a list of required reports and their submittal frequencies for input into the simulation model for resource prediction of future contracts.

Since report preparation times were not recorded by the project team, the minimum, most likely, and maximum values for the preparation time of each report were provided by company experts based on prior experience. Individual labor rates for each personnel type were not provided by the industrial partner; therefore, an average labor rate of 60 CAD per hour was input into the model. A subset of the data is summarized in Table 2.5.

Table 2.5: Sample of report preparation-associated input data

Report Name	**Frequency**	**Resources**	**Time (Minutes)** ***Min, Most Likely, Max***
Daily Update: work plan and estimated progress	Daily	1 Safety Coordinator 1 Project Manager 4 Superintendents 1 Quality Controller	45, 60, 75
Equipment Log	Bi-weekly	1 Project Coordinator	90, 120, 150
Installation Work Package Report	Bi-weekly	1 Project Controller 1 Scheduler 1 Project Manager	210, 240, 270

3-Week Look-Ahead Schedule	Bi-weekly	1 Project Control 1 Scheduler 1 Project Manager	360, 420, 480

2.6.2 Results and Discussion

2.6.2.1 *Extraction Module*

A sample of the extracted reporting requirements is illustrated in Figure 2.3. As detailed in Section 2.6.1, 340 individual reporting requirements and their submittal frequencies were identified from over 500 contract pages. Although quite high (88%, Table 2.4), the recall of the current extraction module is not 100%. Sufficient for planning purposes, the extraction module should not be used as the only means of requirement extraction during the execution phase of a project. A manual review during the execution phase should continue to be performed until the ability of the framework to consistently extract 100% of reporting requirements is achieved. Failing to determine the exhaustive list of submittals and information deliverables required by the client can result in claims and litigations, subjecting both parties to disputes and conflicts that could have been prevented. Nevertheless, manual extraction is also prone to error, and the use of the extraction module as an adjunct tool during the execution phase of a project is strongly recommended. The list of reporting requirements extracted manually and by the automated extraction module should be compared to identify requirements that may be missing from the manually extracted list.

2.6.2.2 Prediction Module

The Monte Carlo simulation model was run for 100,000 iterations, as increasing the number iterations beyond 100,000 slowed the execution speed without resulting in a notable impact on output results. The total duration and cost associated with the requirement reporting process was calculated using the probability distributions defined for each report. In each iteration, random numbers were sampled from the preparation time distributions of each report type, and a total reporting duration (or total cost) was achieved as the cumulative time (or cost) of all reports for that iteration. Total reporting duration (or cost) values of each iteration were then used to form a distribution of total reporting time, as shown in Figure 2.11.

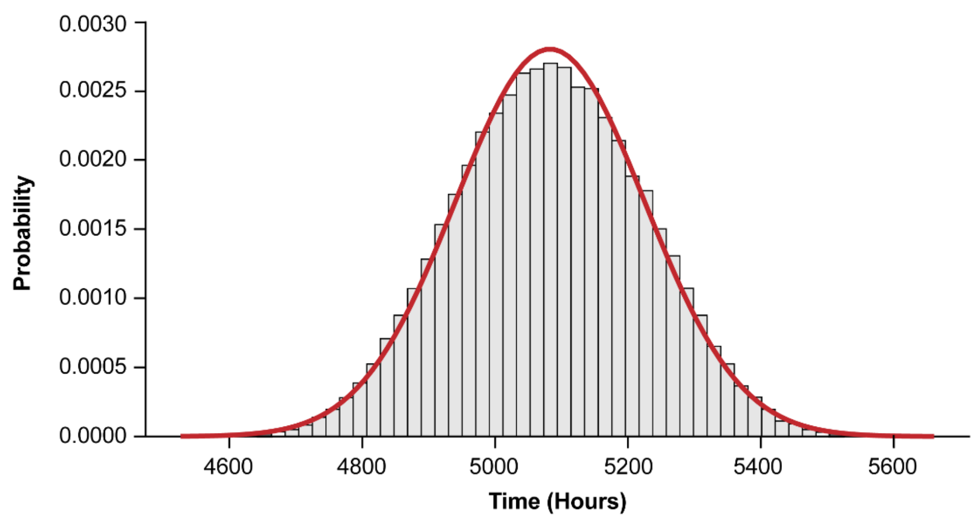


Figure 2.11: Predicted cumulative report preparation time as a distribution

The mean value for the cumulative report preparation time was 5083 labor-hours ($\sigma = 142$) for the project life cycle. The simulation was then run again using an average rate of 60 CAD per hour; here, the mean value of the total cost associated with the reporting process was calculated to be \$304,939 (σ = \$8538), as shown in the predicted cost distribution in Figure 2.12. Notably, including individual labor rates for each personnel type will increase the accuracy of the framework's results.

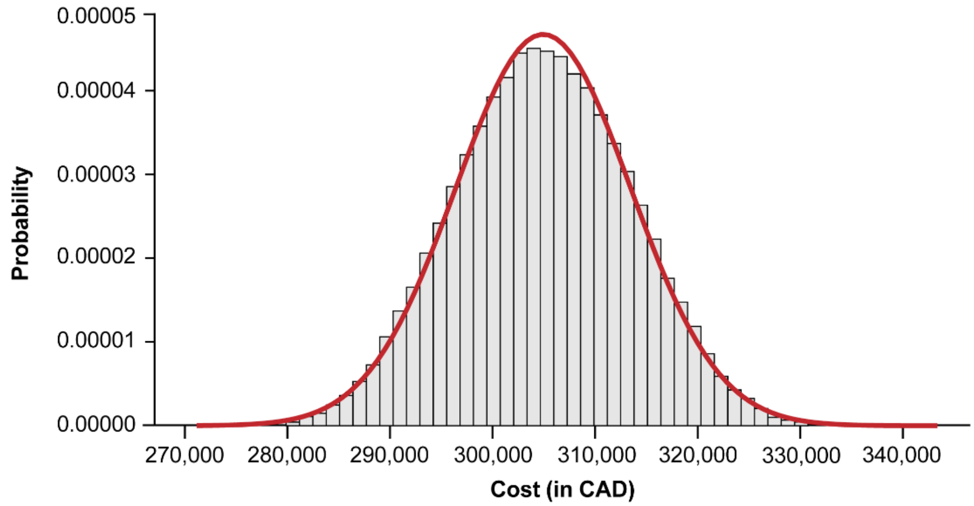


Figure 2.12: Predicted cumulative cost of reporting

The distribution of report preparation time between various personnel types is summarized in Figure 2.13. The plurality of the cumulative report preparation time (31%) was associated with the project manager, who must review and approve most reports. Based on the mean cumulative duration of 5083 h (Figure 2.11), the project manager is expected to spend an estimated 1576 h (or, assuming a 9-h work day, 175 days) completing reporting requirements. With a provided project duration of 4400 h, the project manager is estimated to be performing reporting activities 36% of the time. Similarly, two other highly utilized resources were the project control team (28%) and scheduler (28%), who are responsible for collecting, merging, and overseeing the preparation of various report types. Together, these two resources will spend an estimated 2846 h (or 316 days) completing reporting requirements—equal to 32% of their time.

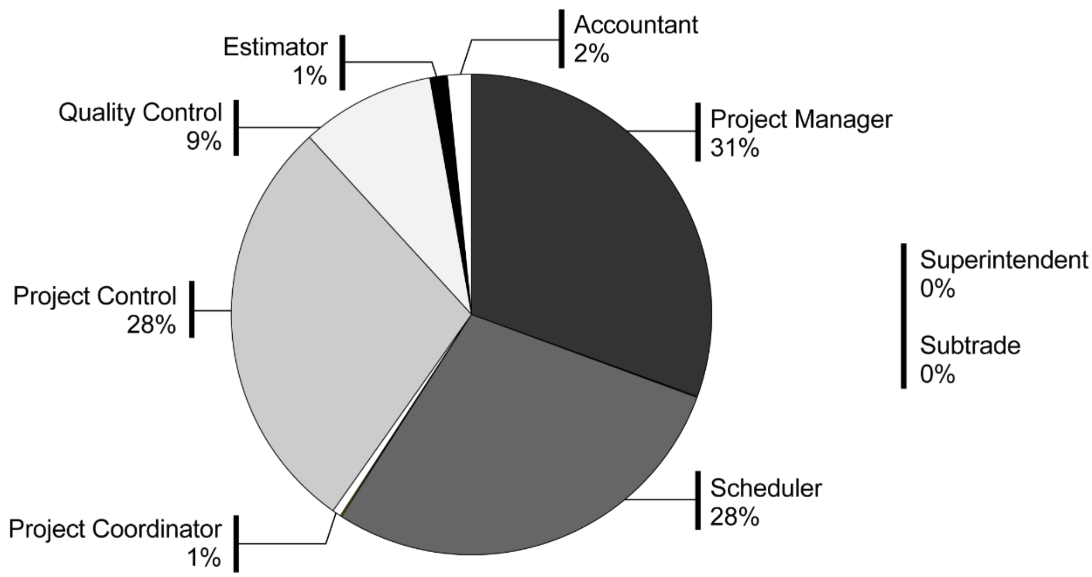


Figure 2.13: Distribution of report preparation time between personnel types

2.6.3 Framework Validation

2.6.3.1 *Validation of Extraction Module*

The extraction module underwent extensive validation testing. Here, reporting requirements were manually extracted and compared to the list of reporting requirements identified using the extraction module. Then, a subset of real project data (different from those used for model development and training) was used to evaluate the performance of the rule-based and machine-learning-based classification models. A discussion of the validation process is detailed in Sections 2.4.2.1 and 2.4.3.1, respectively. A comparison of the models is summarized in Section 2.4.4.

2.6.3.2 *Validation of Prediction Module*

In contrast, the prediction module was evaluated using face validation. Since actual report preparation durations were not recorded by the contractor, face validation was performed by subject-matter experts to evaluate whether or not the simulation model results (i.e., prediction module outcomes) were accurately representing the current reporting process. Simulation results (Figure 2.11-Figure 2.13) were presented to the project management team responsible for executing the case study project. The experts confirmed that the simulated results were acceptable and were consistent with the outcomes of the actual project. Overall, face validation by the subject-matter experts confirmed that the prediction module was representative, comprehensive, and easy to use.

2.7 Discussion

Having a list of reporting requirements during the planning phase of a project will provide the project management team with the opportunity to enhance the reporting process, resulting in a reduction in reporting-associated costs. For example, similar or redundant reports can be consolidated, specialized data collection systems and report templates can be developed and implemented prior to project execution, and the allocation of reporting requirements to specific personnel types can be optimized.

The probability distributions output by the proposed framework allow decision-makers to more accurately estimate the probability of achieving project targets, while gaining insight on potential best- and worst-case scenarios. More accurate time preparation estimates will allow contractors to ensure that a sufficient number of personnel are available to complete reporting requirements on time. Moreover, by more accurately estimating the overhead costs associated with reporting requirements for each particular project, contractors are able to enhance bid preparation to improve competitiveness, or provide more realistic direct–indirect cost ratios to avoid potential disputes for cost-plus contracts. Furthermore, these outputs can be used to optimize the composition of project management teams based on the specific requirements of each contract. Together with the list of requirements output from the extraction module, the personnel distribution results can be used to examine and potentially reallocate reporting duties to lower-wage personnel, where appropriate, thereby reducing report preparation costs.

The level of benefit achieved by considering reporting-associated costs in the planning phase of construction depends on the construction type. Repetitive types of construction, such as residential building construction, are typically associated with contracts that remain similar between projects. Due to a lack of variability in reporting requirements, project teams are able to accurately approximate the time, cost, and resources required without the need to extract reporting requirements for each contract. However, due to the increased level of risk, complexity of the work, and large project scale, contract documents for complex projects, such as those in the oil and gas industry, are typically longer, more intricate, and more variable from project to project. With these types of construction, clients tend to request additional information and detailed reporting submittals from contractors, which substantially increases overhead costs. The benefits of applying the proposed framework, therefore, are expected to expand as project complexity is increased.

2.8 Limitations and Future Work

An automated approach for rapidly extracting reporting requirements from contractual documents and predicting the time and cost required to complete reporting activities was developed. Although the functionality of the proposed framework was demonstrated using real contractual documents from an actual case study, the following points should be considered.

First, the extraction module was developed using a labeled dataset obtained from one set of contract and specification documents for an oil and gas project. While the extraction module is expected to be applicable—in its current form—to all construction contracts with similar characteristics (e.g., terminology, document structure, and/or report structure), the development methodology described may need to be reapplied for other contract types. Moreover, the classification models were trained using a limited amount of training data. The comparatively low performance of the classification models for the "true" class may be due to the size of the "true" dataset (i.e., an imbalanced data problem). Future work should examine the impact of increasing the training dataset through the incorporation of additional contract documents to enhance the performance of the classification models. With sufficient training data, the extraction module is anticipated to achieve the desired performance of 100% recall for the "true" class (i.e., identification of all reporting requirements).

Second, the success of the prediction module is highly dependent on accurately modeling the inputs. One of the difficulties in analyzing probabilistic processes inherent to construction is defining the probability distributions that best reflect the uncertainties associated with each variable. The more accurate the model of the inputs, the more closely the simulation model mimics real-life behavior. A primary constraint for any simulation model, therefore, is the time and effort required to collect pertinent and correct information, as well as processing it for input into the model. While the resources and time required from construction sites and administration offices to complete reporting requirements are not commonly recorded, efforts to improve data collection related to project reporting process are expected to improve model results.

Third, contract documentation remains an immature area of practice, and more reliable and efficient approaches to better and more rapidly understand contract requirements are needed. Future work should focus on providing a holistic solution to this problem, such as writing contracts using a structured-database approach. While this would provide seamless integration between clients and contractors (thereby alleviating the need for rule-based/ML-based model (re)training), achieving this ideal will require a tremendous amount of input, effort, and collaboration among all of the stakeholders involved in a project. Additionally, methods for dealing with modifications or alternate arrangements will need to be researched and developed. Consequently, the framework proposed here provides a much-needed interim solution as these more holistic solutions are pursued.

2.9 Conclusion

Automating the reporting requirement extraction process and estimating its associated time–cost implications are expected to reduce the effort, time, and overhead costs expended by the multiple personnel involved. To overcome the shortcomings of the traditional manual approach, this study developed a framework for reporting requirement extraction based on NLP—a domain-specific and application-oriented text classification process—that is capable of automatically identifying reporting requirements from contractual documents to considerably reduce the time and effort required to extract reporting requirements. To account for project uncertainties due to variation or unforeseeable events that may occur during execution, a Monte Carlo simulation was used to predict the time and cost needed to complete reporting requirements.

The model begins by collecting textual data, in this case the sentences and terms in contractual documents, which describe the reporting requirements mandated by the client. Rule-based and ML-based classification methods were developed, and their performances were evaluated. The performance of rule-based classification using different sets of n-grams was assessed, with an accuracy of 99.27% achieved using bi-grams as rules. Application of lemmatization to and removal of stop-words from the bi-gram rules resulted in a recall and F1-score of 88% and 92% for the "true" category, respectively. Four ML algorithms were also implemented, and their performance was assessed under different pre-processing settings and feature engineering techniques. All of the ML classification models achieved promising accuracies of over 95%; notably, XGBOOST achieved the highest recall value of 89% after parameter tuning. Then, numerical data regarding report preparation times and associated resources (based on prior experience of experts) were provided by an industrial partner and were used to predict the time and cost required to complete the reporting requirements detailed in the contractual documents. Input of these data into the Monte Carlo simulation model resulted in a mean cumulative reporting duration and cost of 5083 hrs and 304,939 CAD, respectively.

During the bidding and contract negotiation phase of a project, decision-makers can now use the proposed framework to automatically review reporting requirements prior to accepting the contractual agreement. Not feasible using time-consuming, traditional extraction methods, the extraction speed of the framework allows decision-makers to identify and subsequently negotiate difficult and/or inefficient reporting requirements prior to signing. If contract conditions are unfavorable to the contractor in terms of project reporting cost, a revision of contract conditions may be requested or a contract may be abandoned by contractors to prevent further loss. With a thorough and realistic understanding of contract reporting requirements, contractors can focus on establishing the best means, methods, pricing, and schedules for completing the proposed project.

2.10 Acknowledgments

This research work is funded by a Collaborative Research and Development Grant (CRDPJ 492657) from the Natural Science and Engineering Research Council. The authors would like to thank Graham Industrial Services LP for their support and for providing contractual documents and report preparation-associated data. The authors also would like to acknowledge Dr. Catherine Pretzlaw for her assistance with manuscript editing and composition.

Chapter 3

Integrating Value Stream Mapping and Discrete-Event Simulation to Improve Administrative Processes Within the Construction Industry

3.1 Introduction

In construction, two different types of processes exist (Belayutham *et al.*, 2016): the production processes that generate a visible output and the administrative processes that support the core production activities of a construction project (Alves *et al.*, 2016). Administrative processes interconnect numerous activities onsite to generate the information required for production processes to occur. These activities, which include progress reporting, procurement assignments, and financial tasks, act a central flow point for many production-related activities (Alves *et al.*, 2016). Expectedly, information loss and interruption resulting from administrative inefficiency and uncertainty (Belayutham *et al.*, 2016) has been identified as a primary cause of delay in the construction industry (AlSehaimi and Koskela, 2008).

Administrative processes represent a considerable portion of project budgets, accounting for up to 80% of the costs incurred to meet client requirements (Jafari *et al.*, 2021; Monteiro *et al.*, 2017; Tapping and Shuker, 2003). Yet, in spite of the tremendous time and resources involved, it has been estimated that only 1% of administrative efforts are value-adding, with approximately 99% (49% non-value-adding and 50% supporting but non-value-adding) considered waste (Hines and Taylor, 2000). Lean construction is quickly emerging as an effective means of improving project delivery by minimizing waste (Koskela *et al.*, 2002). While the benefits of Lean construction are numerous and well-reported (Babalola *et al.*, 2019), they have traditionally focused on improving production processes. Indeed, after a thorough review of construction literature, Yokoyama and colleagues reported that the application of tools and Lean concepts to improve administrative processes in construction was examined by less than a dozen studies (Yokoyama *et al.*, 2019).

Despite the small number of studies, however, the application of Lean concepts to increase administrative efficiency, also referred to as Lean office, is anticipated to result in substantial improvements to construction project delivery. Lean office has been shown to improve work flow, productivity, teamwork, client satisfaction, quality of service, cycle times, and work place organization while reducing lead times, rework, and project costs (Monteiro *et al.*, 2015; Naftanaila and Mocanu, 2014; Tapping and Shuker, 2003). While these benefits have been observed across several domains, including manufacturing (Chen and Cox, 2012), health services, education (da Silva *et al.*, 2015), and electric, gas, and sanitary services (Monteiro *et al.*, 2015), the impact of applying Lean office in construction administration remains relatively unexplored. Indeed, a Lean concept-based method for identifying potential areas of improvement in construction administration, as well as a method for evaluating the potential impact of these improvements on project outcomes, has yet to be developed.

To address this gap, this research is proposing a well-defined, reproducible approach for identifying areas of improvement in construction administration and quantitatively predicting their impact on project performance for enhanced decision-making. Using Value Stream Mapping

(VSM) and Discrete-Event Simulation (DES), the framework stochastically models administrative processes and uses performance metrics to (1) quantitatively assess the performance of current administrative processes, (2) identify potential improvements to current processes, and (3) quantitatively predict the impact of proposed improvements on future project performance. Applicability of the DES and Lean concept-based approach was demonstrated through a case study of a real project, where the proposed framework and metrics were found capable of reliably quantifying current and predicting future administrative performance of a construction contractor.

Highlighting the need for fundamental approaches that promote administrative efficiency, this work is expected to not only facilitate the elimination of administrative waste but to also justify and promote the adoption of Lean decision-making in construction. Importantly, as the first reported method for quantifying and predicting administrative performance in construction, this study is anticipated to serve as a foundation for the development of future Lean office strategies and tools in the construction industry.

3.2 Research Background

3.2.1 Lean Thinking and Lean Construction

Originating as a term to describe the Toyota Production System, Lean thinking has emerged as a philosophy centered around increasing value and minimizing waste in business processes (Hamzeh, 2011). Lean concepts focus on increasing value through continuous improvement by building a culture of teamwork, maximizing flow, and reducing activities or processes that do not add value (i.e., waste) (Liker, 2004; Ohno, 1988; Womak *et al.*, 1990). Lean philosophy was initially adopted by the manufacturing sector across a variety of industrial settings. In 1992, Koskela (Koskela, 1992) reviewed the foundations of what he called New Production Philosophy and proposed a number of modifications that would render Lean thinking more suitable for construction (Koskela, 1992). Since this time, the application of Lean concepts in construction, often referred to as Lean construction, has resulted in outstanding improvements in the performance of countless construction projects (Ballard and Howell, 2003; Howell, 1999). Indeed, a systematic review reported 20 different economic, social, and environmental benefits stemming from the application of Lean concepts in construction (Babalola *et al.*, 2019), including improved health and safety; reduced lead times, delays, and costs; and increased transparency, profit reliability, quality, and customer satisfaction (Ghosh and Burghart, 2019).

Application of Lean philosophy explores process activities from three perspectives: transformation, flow, and value. Transformation separates an entire process into pieces, where each piece is treated as independent, yet also connected to other pieces by way of outputs and inputs (Tommelein, 2015). Flow examines the resources and buffers (i.e., connections between the pieces) required to transform the pieces. The value perspective aims to understand and deliver what customers expect.

Process activities can be classified into three different categories (Tyagi *et al.*, 2015). The first, value-adding (VA), are activities that push the process forward, creating value for clients. Second, necessary but non-value-adding (i.e., necessary waste, NW), are supporting activities that do not add value to the output but are necessary to ensure the flow of the value-adding activities

(Belayutham *et al.*, 2016). Finally, non-value-adding (i.e., pure waste, PW), are activities that do not move the process forward and provide no value for the client. Lean philosophy focuses on identifying and eliminating non-value-adding activities (i.e., NW and PW) to improve value and increase flow.

3.2.2 Lean Office in Construction

In 2003, Tapping and Shuker adapted Lean principles to suit office and administrative activities to improve these processes (Tapping and Shuker, 2003). Since this time, many Lean principles have been modified and applied to office environments to improve administrative processes in a practice known as Lean office (East and Love, 2011; Tapping and Shuker, 2003). Later, Hicks promoted the application of Lean thinking to support the improvement of information management and information systems infrastructure (Hicks, 2007). Lean office strategies have primarily focused on information flow and employee knowledge (Monteiro *et al.*, 2017), since the flow of information is necessary to support the flow of production-related activities (Kemmer *et al.*, 2009).

Administrative processes in construction concentrate on managing the flow of project information and documents, such as reports, change orders, contracts, invoices, and payments (Alves *et al.*, 2016; Jafari *et al.*, 2020). Efficient flow of information and documents is an essential component of successful project management (Kemmer *et al.*, 2009), with bottlenecks in information flow estimated to result in up to 37% of non-value-adding time in construction (Al Hattab and Hamzeh, 2018). Due to production waste appearing more tangible and visible, existing Lean construction literature has focused on the application of Lean principles to production processes (Belayutham *et al.*, 2016; Keyte and Locher, 2004). Waste in administrative processes is less apparent than in production and—due to greater variation in office processes, the absence of a visible physical flow, and a lack of directives or foundations—is more difficult to minimize (Chen and Cox, 2012).

Consequently, very few studies have examined or developed strategies for improving the management of office and administrative processes in construction (Costa *et al.*, 2013; East and Love, 2011; Garrett and Lee, 2010; Kemmer *et al.*, 2009; Ko and Li, 2015; Lima *et al.*, 2010; Pestana and Alves, 2012; Rossiti *et al.*, 2016). A summary of the findings is detailed in Table 3.1. As shown in Table 3.1, many of the previous studies exploring Lean office in construction were limited to solving a specific subset of administrative processes of one specific company (i.e., a case study). Frameworks that could be implemented in other administrative problems were not developed, and the performance metrics used in the previous studies were limited—as a result of the specificity of the research—to a particular administrative process. Importantly, due to a lack of suitable methods, these studies were not able to thoroughly assess and predict the quantitative impact of proposed improvements.

Table 3.1: Summary of previous studies on Lean office in construction

Research Study	**Type of Study**	**Application Area**	**Performance Metrics**
Lean office at a construction company (Kemmer et al., 2009)	Case Study	Billing and Payment	• Problem found • Non-authorized RFPS
Lean construction submittal process—A case study (Garrett and Lee, 2010)	Case Study	Construction Submittals Review	• Process time • Lead time • Number of activities
Value stream mapping of the architectural executive design in a governmental organization (Lima et al., 2010)	Case Study	Architectural Executive Design	• Process time • Lead time
Value-added analysis of the construction submittal process (East and Love, 2011)	VA Assessment Framework	Submittal Preparation Process	• Process duration
Mapping the submittal process in a design-bid-build project (Pestana and Alves, 2012)	Case Study	Submittal Process in a Design-Bid-Build project	• No performance assessment
Redesigning administrative procedures using value stream mapping: A case study (Costa et al., 2013)	Case Study	Buying and Suppliers' Payment Process	• Number of employees • Total time • Process time
Lean concurrent submittal review systems (Ko and Li, 2015)	Lean Concurrent Submittal Review System Framework	Submittal Review	• Approval rate • Number of change orders • Review time
Impacts of Lean office application in the supply sector of a construction company (Rossiti et al., 2016)	Case Study	Supply Sector	• Total cycle time • Lead time • Added value

3.2.3 VSM Integration with Simulation

Lean concepts use mapping tools to analyze and evaluate current systems (Rother and Shook, 2003). One of the most powerful and commonly-applied tools is Value Stream Mapping (VSM) (Tapping and Shuker, 2003), which is used to map all of the activities necessary (both value-adding and non-value-adding) to deliver a finished product or output to the client (Rother and Shook, 2003). Once the process is mapped, activities in the value stream are classified by value-adding type, and actions for minimizing waste are identified and applied. While VSM has been found capable of improving a number of project outcomes, including productivity, process visibility, cost, labor hours, lead time, and processing time (Shou *et al.*, 2017), several shortcomings have been attributed to VSM. Specifically, VSM is limited to the creation of a static model, cannot consider inherent variability, and is unable to analytically evaluate the process (Atieh *et al.*, 2016; Luz *et al.*, 2020).

These limitations prevent practitioners from evaluating the impact of proposed changes before implementation—particularly in consideration of the uncertainty that is inherent to construction. Researchers have coupled simulation with VSM to overcome the rigid and deterministic nature of VSM. Hybrid simulation-VSM has enabled the analysis of data in a dynamic environment, the evaluation of changes before implementation, and the quantification of improvements during planning and assessment phase (Abdulmalek and Rajgopal, 2007; Erikshammar *et al.*, 2013). While hybrid simulation-VSM approaches have been used to improve a number of production-related processes in various domains such as the manufacturing sector, a handful of studies have applied simulation-VSM-based approaches in construction (Erikshammar *et al.*, 2013; Fontanini *et al.*, 2008; Wang *et al.*, 2009; Yu *et al.*, 2009; Zahraee *et al.*, 2021). Details of these studies are listed in Table 3.2. These studies focused on the production process of various domains, neglecting construction administration. To the best of the authors' knowledge, hybrid simulation-VSM has not yet been applied to administrative processes.

Table 3.2: Summary of previous studies using simulation-VSM-based approaches in construction

Research Study	Process Type	Simulation Type	Application Area
Simulating a construction supply chain: Preliminary case study of pre-cast elements (Fontanini et al., 2008)	Production	System Dynamics Simulation	Supply chain for pre-cast concrete elements
Development of Lean model for house construction using value stream mapping (Yu et al., 2009)	Production	Discrete Event Simulation	Housing construction process
Flow production of pipe spool fabrication: simulation to support implementation of Lean technique (Wang et al., 2009)	Production	Discrete Event Simulation	Pipe spool fabrication process
Discrete event simulation enhanced value stream mapping: An industrialized construction case study (Erikshammar et al., 2013)	Production	Discrete Event Simulation	Patio doors production process
Lean construction analysis of concrete pouring process using value stream mapping and arena based simulation model (Zahraee et al., 2021)	Production	Discrete Event Simulation	Concrete pouring process

3.2.4 Research Gap

A considerable potential for improvement exists in the area of construction administration, with benefits expected to be as impactful as those achieved in production (Monteiro *et al.*, 2015). Yet in spite of their importance, the management of administrative activities has often been overlooked by both researchers and practitioners. Previous studies have focused on identifying and classifying Lean tools and techniques, with very little effort focused on systematically identifying and categorizing various Lean construction practices and their benefits. Although hybrid simulation-VSM approaches have the potential to address these gaps, the application of simulation-VSM-based methods is still in its infancy and is not systematically integrated. Reproducible methods capable of identifying areas of improvement and quantifying and predicting administrative performance in construction have yet to be developed.

3.3 Methodology

To address these limitations, this research has developed a hybrid VSM-discrete event simulation (DES)-based solution that can be used to identify areas of improvement in administrative processes and quantitatively predict their impact on project performance. The proposed solution was developed using the Design Science Research (DSR)-based process model proposed by Peffers and colleagues in 2007 (Peffers *et al.*, 2007) as follows.

Problem definition and solution objectives for this research were identified following the review of academic literature and through discussions with experienced professionals. The literature search focused on identifying research gaps and limitations in the areas of construction engineering and management, Lean construction, and Lean office. Informal discussions were held with several industrial practitioners currently working in the construction sector in Alberta, Canada. Desired objectives of this research were to (1) improve administrative efficiency in construction, (2) establish performance indicators in construction administration, and (3) develop a method capable of quantifying proposed improvements. An artifact for achieving the desired objectives (herein referred to as the framework) was developed. A case study of a real construction project was used to demonstrate the ability of the framework to solve an instance of the problem and to evaluate how well the framework was able to achieve a solution to the problem. Outcomes of the research, including conclusions, limitations, and future work, were examined and communicated.

3.4 Proposed Framework

The proposed framework, summarized in Figure 3.1, is comprised of three components:

1. *Measurement and Assessment*, which involves understanding and measuring the current administrative process by visualizing the process' attributes.

2. *Analysis and Improvement,* which involves identifying root causes of waste in the process and proposing opportunities for improvement based on the findings of the previous step.

3. *Performance Prediction and Implementation*, which involves experimenting with proposed improvements and estimating the resulting impact on future performance.

Lean philosophy is applied throughout the framework to measure, identify, and implement opportunities for improvements. To consider the uncertainties inherent to construction, DES is used to assess current and predict future performance metrics and their variability. The proposed framework can be re-applied following the implementation of proposed improvements to re-assess process performance, thereby facilitating and promoting a culture of continuous improvement in construction administration.

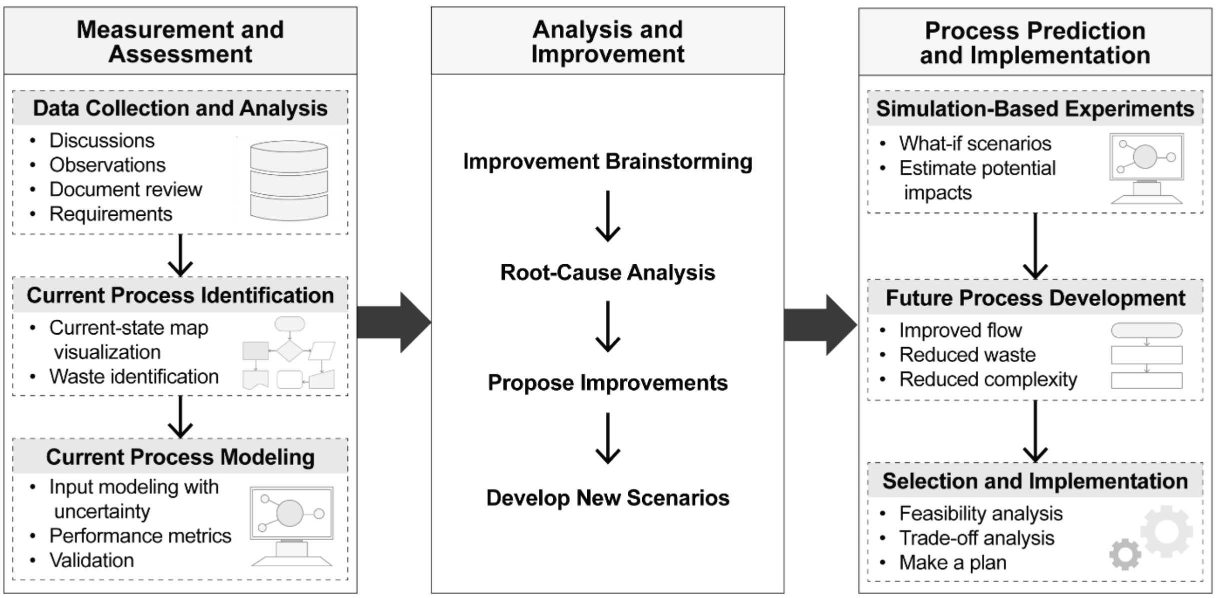


Figure 3.1: Overview of the proposed framework

3.4.1 Process Assessment and Measurement

This component measures and evaluates the performance of the current process using performance metrics thoughtfully selected by each organization to reduce time, rework, and complexity. Quantitative input data describing the current process are collected, and VSM is integrated with DES to portray, simulate, and evaluate the performance of the administrative process. Results of the assessment and measurement component will provide a baseline to evaluate the magnitude of future process improvements.

3.4.1.1 Data Collection and Analysis

First, a list of current administrative tasks is prepared using information collected from various sources (e.g., contract documents, corporate database systems, subject matter experts) using a variety of methods (e.g., interviews, observations, and review of documentation), as appropriate. The frequency, duration, and the resources required to complete the task, as well as the likelihood and requestor of administrative rework, are also collected. Rework in construction administration takes the form of document revisions, which can be classified as (1) client-based, including all revisions requested by the client prior to approving the document and (2) internal, including all revisions requested internally by members of the project team that are responsible for reviewing and approving the document before submission to the client.

Task durations can be collected as either discrete values or as probability distributions. Data requirements and collection methods should be established in parallel with current process mapping and simulation modeling to define system layout, parameters, and operating procedures. As a critical step of the framework, omissions or errors in this step may nullify further analysis.

Data collection and analysis may be a challenging step when applying Lean office in practice, as data associated with administrative processes in construction are often not readily available or in a proper format. However, if data are unavailable, estimates can be used. For example, task durations that are not recorded in a database can be estimated by personnel with knowledge of the process. Similarly, event occurrences, such as rework, can be estimated by knowledgeable personnel as likelihoods (i.e., the number of times it has occurred in previous projects), in consideration of factors such as project size and complexity, client requirements, and the number of sub-trades.

3.4.1.2 Current Process Identification

Then, VSM is used to develop a value stream map that details the current state of the administrative process under study. Here, the analyst observes and explores the system and, using Lean techniques, draws an as-is (i.e., current) state map to record the process. The current-state value stream map is used to identify waste in the value stream. Activities are classified as value-adding (VA), necessary waste (NW), or pure waste (PW) based on the nature of the administrative work, as detailed by Tyagi and colleagues (Tyagi *et al.*, 2015).

3.4.1.3 Current Process Modeling

To better represent process variability and to allow for quantitative evaluation, a DES model is then built from the value stream map. DES has been proven capable of analyzing, evaluating, and predicting behaviors of a system before implementation (AbouRizk, 2010) while handling uncertainty and generating performance statistics to support continuous improvement (Wang *et al.*, 2009). Using DES, the administrative process can be evaluated and re-designed to achieve the desired performance values. As such, DES was chosen as the preferred simulation method for this framework as it is well-suited to either complement or substitute VSM (Goienetxea Uriarte *et al.*, 2020).

In proposed approach, both managerial and operational aspects of administrative processes can be considered. The developed simulation model is capable of generating performance statistics and resource requirements while remaining flexible to specific process details. To consider uncertainty, durations are modeled as probabilistic distributions. These can be derived from historical data or, if data are unavailable, can be generated from expert opinion, using minimum, maximum, and most likely values.

Once built, the current-state value stream map containing collected data, including activity durations, assigned resources, rework likelihoods, and anticipated project duration, are input into the simulation model. The DES approach incorporates uncertainty and dynamicity by allowing durations to be modeled as probabilistic distributions, such as uniform distributions with minimum and maximum values or triangular distribution with minimum, most likely, and maximum values. The resulting model is capable of generating performance statistics and resource requirements, while remaining flexible to specific process details. Performance metrics, together with the simulation model, are then used to quantitatively evaluate the current process. A list of performance metrics applicable to many construction administrative processes are detailed in Table 3.3. The number and/or definitions of the performance metrics should be adjusted to suit the specific needs of an organization.

Table 3.3: Recommended performance metrics

Performance Metric[1]	Description
Time Total	Time to complete all administrative activities
Time VA	Time to complete all value-adding activities
Time NW	Time to complete all necessary waste activities
Time PW	Time to complete all pure waste activities
Rework Total	Time to perform rework
Rework Client	Time to perform rework requested by client
Rework Internal	Time to perform rework requested by project team
Resource Total	Time required by each resource type
Resource VA	Time required by each resource type on value-adding activities
Resource NW	Time required by each resource type on necessary waste activities
Resource PW	Time required by each resource type on pure waste activities
Steps Total	Number of process steps
Steps VA	Number of VA process steps

[1]VA: value-adding, NW: necessary waste, PW: pure waste

The developed simulation model is able to handle uncertainty and create dynamic views of the selected measurement metrics, such as durations for each activity, processing times, resource time, prediction of the amount of rework in the process, and the total volume of expected work. Verification of the model should be performed concurrently with model development to determine whether or not the model is mimicking the real system and to identify undesirable system behavior (Sargent, 2010). Once built, the model should be further validated by comparing simulation results to the actual process. If historical project data are not available, the face validation technique may instead be used (Sargent, 2010).

3.4.2 Process Analysis and Improvement

Completion of the process measurement and assessment component will result in the generation of performance values across a variety of performance metrics for the current process. Using these results, the process analysis and improvement component of the proposed framework is initiated. Here, improvements designed to minimize time spent on non-value-adding activities through the reduction or elimination of root causes are identified.

Brainstorming is a commonly-applied and highly-effective tool used to generate creative solutions (Tyagi *et al.*, 2015). This technique capitalizes on the diverse experience of all team members involved in the administrative process to examine the problem from a variety of perspectives and from novel angles. Experts in various areas, such as project management, contract administration, and data management systems, should be included to ensure that all root causes—particularly those specific to the organization—are identified. Once root causes are determined, potential improvements that can be feasibly implemented at the organization are proposed.

Then, based on the proposed improvements, new experimental state value stream maps are developed. Depending on the number of improvements proposed, multiple experimental state maps representing a number of scenarios may be developed, particularly when conflicts preventing multiple improvements from being simultaneously implemented may arise. Experimental maps may include only one or multiple improvements. The experimental value stream maps will include process modifications, such as changes to the duration of activities and the mapped workflow (i.e., addition, removal, and integration of activities).

3.4.3 Performance Prediction and Implementation

This component focuses on predicting future performance of the proposed improvements using a DES-based approach. Similar to the quantitative evaluation of the current-state map, experimental value stream maps are used to build DES models that predict the future performance of each scenario. Using the simulation results, analysts can compare the benefits of the various scenarios and balance these benefits with the level of difficulty, resources, and time required to implement the proposed improvements.

3.4.3.1 Simulation-Based Experiments

Each experimental state map or scenario are run by changing (1) parameter values (e.g., duration estimates), (2) the workflow of the model (e.g., adding, removing, or integrating activities), or (3) both. Resulting distributions for each performance metric are used to predict and evaluate the potential future performance of each experimental scenario.

3.4.3.2 Future Process Development

Performance results for each scenario are compared to determine if any of the experimental alternatives outperform the current process. Ideal scenarios will result in the elimination of non-value adding activities (i.e., PW) and the reduction of necessary but non-value adding activities (i.e., NW). A future-state map is then created based on the experimental scenario(s), resulting in improved performance outcomes.

3.4.3.3 Selection and Implementation

While an ideal scenario may be observed, the feasibility of implementation must also be considered when developing an implementation plan. Implementation should be iterative and incremental, as the complete implementation of an entire Lean system all at once is likely not possible in practice.

Results from the current and experimental states are compared, and options for process improvement can be considered by the organization based on their quantitative impact. The feasibility and cost to implement each of the various scenarios are balanced with the corresponding quantitative impact.

The proposed methodology is then brought to the attention of the personnel responsible for maintaining standard procedures so that a detailed implementation plan for the proposed revisions can be established. Together with input from the project management team, a final implementation strategy is selected and is incorporated into current standard operating procedures. Following implementation, the framework is re-applied to identify further improvement opportunities, with the future state of the current iteration becoming the current state of the subsequent iteration.

3.5 Case Study

To demonstrate functionality of the proposed approach, the framework was applied to the administrative process associated with the preparation of project control reports for an oil and gas construction project led by a Canadian contractor in Alberta. The project was considered a medium-size project (approximately 10 million CAD) by the contractor and was awarded by the client to the contractor through a cost-plus contract type. The project followed a shift-based schedule, with each shift consisting of ten 8-hour working days followed by a period of four non-working days. The project was scheduled to be completed in 16 shifts, resulting in an expected project duration of 160 working days (i.e., 16 shifts of 10 working days) or 1,280 working hours (i.e., 160 working days of 8 hours).

The contractor was contractually-required by the client to submit multiple reports that detailed project progress, any potential problems, and necessary measures that must be taken to accomplish the project objectives. Once generated, construction reports were reviewed and approved by a designated team member to ensure report accuracy before submission to the client. After a report was internally approved, it was submitted to the client, which, in turn, reviewed and approved the report. If a report was incomplete or contained errors, a revision (i.e., rework) was requested either internally or by the client.

3.5.1 Process Assessment and Measurement

3.5.1.1 Data Collection and Analysis

A current list of necessary administrative tasks (i.e., reporting requirements) and their frequencies (i.e., submittal requirements) were extracted from text-based contract documents and project specifications. The expected durations, resources required to complete, rework classification (i.e., client-based or internal), task waste type, as well as the likelihood of rework were collected from:

1. Review of historical records, contract and project specifications, project details, and project documents, such as responsibility assignment matrices.
2. Meetings and discussions with the project management team.
3. Observations of the reporting process (i.e., shadowing of the project management team) by the authors of the study.

The mapped tasks identified during the data collection process (as detailed in Appendix A), the task waste type (i.e., VA, NW, or PW), the task frequency, the task duration, and the assigned resources (i.e., team members) are listed in Table 3.4. To incorporate variability and uncertainty, task durations were collected from the project team as either a range (e.g., 8 to 10 hours) or as discrete values (e.g., 2, 3, or 4 hours) and modeled as either uniform or triangular distributions, respectively.

Table 3.4: Administrative process tasks and associated input data

Task	**Name**	**Type**[1]	**Freq.**	**Duration (h)**[2]	**Resources**[3]
1	Enter timesheets into ERP	NW	daily	U (2-3)	PC
2	Timecard keying	NW	daily	U (5-6)	PA
3	Subtrade tracking in spreadsheet	PW	daily	U (4-6)	PC
4	Update equipment timesheet	NW	daily	Tri (0.75, 1, 1.25)	PC
5	Complete material receiving	NW	daily	Tri (1.5, 2, 2.5)	PA
6	Confirm completion of tasks	NW	daily	0.5	PC
7	Quantity entry	NW	weekly	Tri (3, 4, 5)	PC/PCT
8	Review and code subtrade invoices	NW	weekly	U (3-4)	PCT
9	Key subtrade invoices	NW	weekly	Tri (1.5, 2, 2.5)	PA
10	Post equipment timesheet	NW	weekly	Tri (0.75, 1, 1.25)	PC
11	Run LEMS	NW	shift	2	PA
12	Pull raw data from ERP	PW	shift	1	PA
13	Pull LEMS log	NW	shift	1	PA
14	Prepare LEMS summary	VA	shift	Tri (0.75, 1, 1.25)	PA
15	Review and approve LEMS summary	NW	shift	0.5	PM
16	Prepare and review CCO log	VA	shift	Tri (0.5, 1, 1.5)	PCT
17	Pull client report raw data from ERP	PW	shift	1	PCT
18	Prepare construction schedule	VA	shift	Tri (2.5, 3, 4)	SCH
19	Review and update construction schedule	NW	shift	U (2-3)	PM/S
20	Prepare lookahead schedule	VA	shift	Tri (0.5,1, 1.5)	SCH
21	Review and approve lookahead schedule	NW	shift	U (0.75-1)	PM/S
22	Prepare critical path schedule	VA	shift	Tri (0.5,1, 1.5)	SCH

23	Review and approve critical path schedule	NW	shift	U (0.75-1)	PM/S
24	Prepare equipment log report	VA	shift	Tri (1.5, 2, 2.5)	PC
25	Review equipment	PW	shift	U (0.75-1.25)	PC
26	Review and approve equipment log	NW	shift	U (1-1.5)	PM
27	Pull purchase order raw data	PW	shift	1	PCT
28	Prepare purchase order log report	VA	shift	U (1-1.5)	PC
29	Review and approve purchase order log	NW	shift	1	PCT
30	Prepare subtrade log report	VA	shift	2-3	PC
31	Review subtrade log	PW	shift	U (0.75-1.25)	PCT
32	Review and approve subtrade log	NW	Shift	U (1-1.5)	PM
33	Prepare quantities report	VA	shift	Tri (1.5, 2, 2.5)	PC
34	Review quantities report	PW	shift	U (0.75-1.25)	S
35	Review and approve quantities report	NW	shift	U (1-1.5)	PM
36	Compile and prepare shift progress report	VA	shift	U (8-10)	PCT
37	Review and approve shift progress report	NW	shift	U (3-5)	PM
38	Submit bi-weekly reports	VA	shift	2	PCT

[1]VA: value-adding, NW: necessary waste, PW: pure waste
[2]Tri: triangular distribution, U: uniform distribution
[3]PC: project coordinator, PM: project manager, PA: project accountant, PCT: project control, S: superintendent, SCH: scheduler

Likelihood of rework was determined by examining historical data of a similar-sized projects undertaken with the same client. Table 3.5 lists the reports that required revision, the likelihood of revision occurrence, if the revision was requested internally or by the client, and the tasks required to be repeated to address the requested revision. It should be noted that, while the collected data were related to a similar-sized project, tasks durations, resources, and likelihood of rework for the current project may vary depending on specific project features.

Table 3.5: Rework and associated input data

Rework (Revision)	Likelihood (%)	Requestor[1]	Affected Tasks
LEMS Summary Report	25	Client	14, 15
Construction Schedule Report	30	Client	18, 19
Construction Schedule Report 1st Revision	30	PM	18, 19
Construction Schedule Report 2nd Revision	10	PM	18, 19
Equipment Log Report	10	Client	24, 25, 26
Equipment Log Report	15	PM	24, 25, 26
Purchase Order Log Report	15	PCT	28, 29
Subtrade Log Report	90	PCT	30, 31
Subtrade Log Report	50	PM	30, 31, 32
Quantities Report	40	Client	33, 34, 35
Quantities Report 1st Revision	90	S	33, 34
Quantities Report 2nd Revision	40	S	33, 34
Quantities Report	30	PM	33, 34, 35
Shift Progress Report	30	Client	36, 37

[1]PM: project manager, PA: project accountant, PCT: project control, S: superintendent

3.5.1.2 Current Process Identification

After collecting the project information and required input data, current-state value stream maps were prepared and visualized using VSM (as detailed in Appendix B). The reporting processes consisted of 38 process steps. Here, value in the reporting processes was defined as tasks that directly resulted in the fulfilment of reporting requirements by the contractor as mandated by the client. Tasks were classified as VA, therefore, if an intellectual contribution by the assigned resource was required to prepare or submit the report. Of the 38 steps, 11 (29%) were classified as VA, 20 (53%) were classified as NW, and 7 (18%) were classified as PW (Table 3.4). The current-state maps were validated through consensus and face validation by the project team (i.e., subject matter experts).

3.5.1.3 Current Process Modeling

The current-state value stream map, together with the performance metrics detailed in Table 3.3, the activity information in Table 3.4, and the rework information in Table 3.5 were used to develop the DES model. The characteristics of the administrative activities, as well as the mathematical and logical relationships between the components, were formalized, and a DES model was built using Simphony.NET (AbouRizk *et al.*, 2016).

Model Verification and Validation

Verification of the DES model was done concurrently with model development by tracing model entities to determine if the model's relationships were accurate. As administrative performance data were not collected by the company, face validation was performed (Sargent, 2010). The results of the simulation model were reviewed by subject matter experts that had worked on the case study project. Based on the recollection of the study team, the results generated by the simulation model were consistent with what had been observed in practice, thereby validating the accuracy of the simulation model.

Current Process Results

Following verification and validation, the model was run for 100,000 iterations (Figure 3.2). Probability distributions of the durations of the (1) total administrative process undertaken to perform the tasks and (2) duration of VA tasks are illustrated in Figure 3.2a and Figure 3.2b, respectively. A subset of the performance metric results is summarized in Table 3.6.

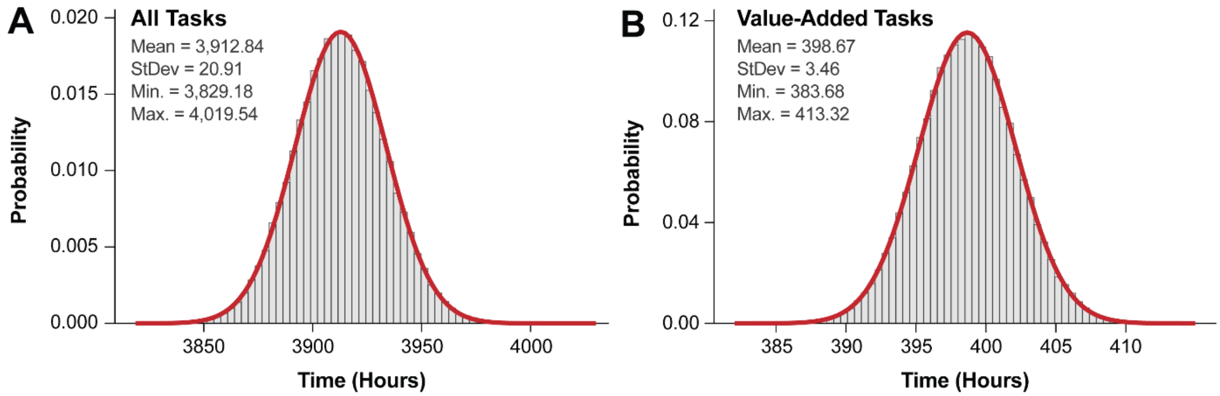


Figure 3.2: Total duration of (A) all tasks and (B) only value-adding tasks of the administration process as probability distributions

Table 3.6: Performance metric results from current state (i.e., baseline) simulation model

Performance Metric[1]	**Simulation Statistics**				**Ratio of Mean**
	Mean	**Std. Dev.**	**Min.**	**Max.**	
Time Total	3912.8	20.9	3829.2	4019.5	-
Time VA	398.7	3.5	383.7	413.3	10.2% of Time Total
Time NW	2443.9	7.3	2412.8	2476.3	62.4% of Time Total
Time PW	1070.2	19	995.9	1159.3	27.4% of Time Total
Rework Total	174.2	17.5	104.2	255.1	16.3% of Time PW
Rework Client	65.8	14	7.3	129.8	38% of Rework Total
Rework Internal	108.4	10.6	63.6	155.2	62% of Rework Total

[1]VA: value-adding, NW: necessary waste, PW: pure waste

In the current administrative process, only 10.2% of the duration was classified as VA, with activities classified as NW and PW accounting for over 90% of the total project duration (Table 3.6). Interestingly, only 16.3% of PW time was attributed to rework, of which 62% was requested internally and 38% was requested by the client (Table 3.6). As expected, time spent engaging in PW tasks was associated with a greater variability compared to VA and NW tasks, as PW tasks include the variability driven by the likelihood of occurrence for rework-associated activities.

The average duration for the total administrative process was 3,913 hours or 489 working days. Given that the project is 160 working days in duration, the project team would require 3 full-time resources to perform all of the tasks in the current-state map. Hours spent by each resource type on VA, NW, and PW activities as a ratio of the total administrative process duration are illustrated in Figure 3.3. The project coordinator was accountable for 44% (1,726 hours) of the total duration—the largest number of hours of all the resource types. The project coordinator was accountable for 82% of PW time, which is due to the nature of the tasks that the project coordinator is assigned to. In contrast, the project controller was associated with the greatest ratio of VA time (4.9% of total duration; Figure 3.3).

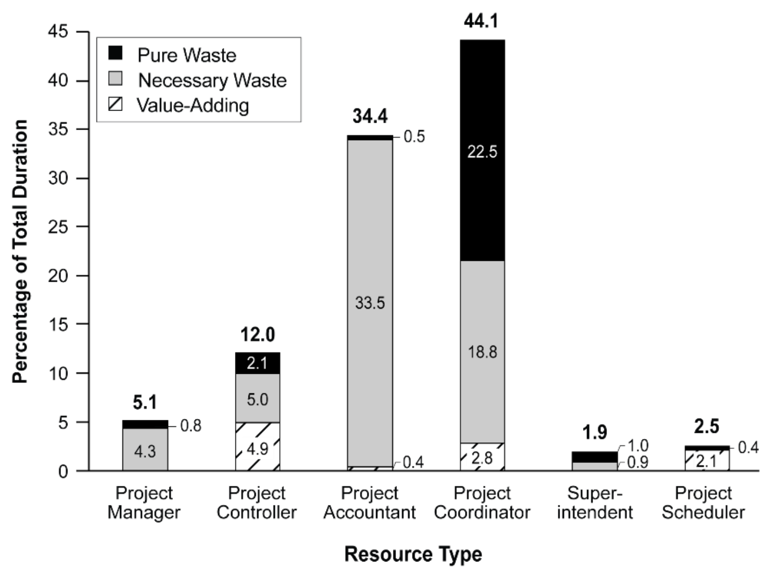


Figure 3.3: Mean value of the time spent on pure waste (black bars), necessary waste (grey bars), and value-adding (hatched bars) tasks by resource type as a percentage of the total duration of the original (i.e., current state) administrative process.

In addition to providing a baseline for the identification and evaluation of future process improvements, the current-state model is able to provide decision-makers with useful decision-support. Separation of tasks by required resource can assist decision-makers with resource planning to ensure a sufficient number and type of staff are available to meet future demands.

3.5.2 Process Analysis and Improvement

Once the baseline performance was assessed using the current-state model, process analysis and improvement was initiated. A brainstorming approach was used by the author and the project team to identify root causes of NW and PW activities.

3.5.2.1 Root Cause Analysis

In the current project, waste was attributed to a number of underlying causes including, but not limited to, (1) multiple rounds of reviews and approvals at various levels, (2) errors during report preparation, (3) manual interaction with existing systems, (4) redundant reporting requirements, and (5) limited availability of and accessibility to up-to-date information. A detailed discussion of the root causes associated with this case study included:

Multiple rounds of reviews and approvals. A number of reports were associated with unnecessary review and verification tasks that did not add value to the report. For instance, after preparation (Task 30), subtrade reports were passed from the project coordinator to project control (Task 31), and then to the project manager (Task 32), for review and approval. In addition to introducing unnecessary steps, multiple rounds of review are unconducive to the Lean principle of Poka Yoke (i.e., mistake-proofing). By creating a sense of security, practitioners are less incentivized to ensure the initial report is free of errors, as it is anticipated that errors will be identified and corrected by those responsible for document review. Furthermore, multiple rounds of reviews can introduce delays. In this case study, certain reports were not passed on immediately upon receipt, thereby delaying the entire administrative process while reports were pending approval.

Errors during report preparation. The first source of errors is related to information exchange. The contractor uses two software systems for reporting: an enterprise resource planning (ERP) system and a separate document management system for generating, distributing, and tracking documents and reports. The use of these two systems results in the fragmentation of project information. The information systems used by the contractor are not able to automatically exchange information with each other, necessitating manual transfer and re-entry of project data. Ensuring that data entry is complete and without errors is of paramount importance, as information incorrectly input or duplicated in the data management system(s) can impact multiple project reports. Another potential source of errors was the lack of training and development of the junior team members (i.e., project coordinators) responsible for preparing a majority of the reports. As the least experienced members, the project coordinators were not as practiced at recognizing anomalies or oversights in reports, reducing the likelihood of identifying mistakes and omissions. Finally, the large workload of the project management team leads to the prioritization of production-related work over report preparation and review. This results in a push system where administrative tasks are often delayed and continue accumulating until immediately prior to the client-mandated submission deadline. The build-up of tasks can defy the capacity of the team member(s), further increasing the defects in the report inventory (i.e., as errors grow exponentially), and prompting further delays.

Manual interactions with existing systems. In addition to introducing errors, as discussed previously, the lack of a comprehensive data management system has resulted in data fragmentation within the organization. Certain project data, such as timesheets and subcontractor data, must be recorded manually in spreadsheets or ad hoc tools and applications. Integration of the information contained within these systems with data from the ERP and document management systems require time-consuming manual reconciliation, in turn increasing the duration of the administrative process.

Redundant reporting requirements. Clients are demanding contractors to provide a growing number of reports as construction projects become increasingly complex. Occasionally, requirements for the various reports mandated by construction contracts and project specification documents are redundant, requiring the duplication of information across multiple reports. For example, the shift progress report (Tasks 36 and 37) includes similar information to many of the other reports, such as the quantities report (Tasks 33,34, and 35). The quantities report must be submitted separately, while the shift progress report also includes quantities information within it.

Limited availability of and accessibility to up-to-date information. While project control reports must be submitted every shift, the project team is generally unable to access up-to-date information required to complete the report from the existing systems. In this particular case study, the project team was required to wait two additional days before data were available in some instances. Also, certain tasks are more complex, requiring considerable effort and internal communication to prevent errors. For example, the project team had to expend considerable time and effort monitoring and accurately controlling the material received on site to generate accurate up-to-date quantities.

Other causes of waste. In addition to what has been previously mentioned, other problems in the current process were identified, including the absence of a formal and standardized reporting process, the failure to identify and manage risk associated with reporting, as well as the failure to recognize the construction reporting process as an activity affecting the total organization (and, therefore, not communicating the issues effectively outside the affected department).

Then, improvements designed to prevent or minimize the impact of the root causes were proposed, and strategies for implementation were developed.

3.5.2.2 Proposed Improvements

Based on the in-depth assessment of the current process, the identification of issues, and the results of the root-cause analysis, potential improvements aimed at eliminating PW tasks while increasing the efficiency of VA and NW tasks were discussed and proposed by the project team and authors of this study following a brainstorming session. Suggested improvements included:

1. Unnecessary requirements for 'review/approval' or multiple rounds of revision should be removed. Instead, review of a report should be completed in one session with all responsible members present.
2. The frequency of certain tasks should be changed from weekly to daily to eliminate the accumulation of errors and decrease the likelihood of rework (i.e., revisions).

3. Process steps can be shortened by combing several steps. For example, rather than having a project coordinator prepare a report, place it aside, and return to the report later for review, the project coordinator should prepare and review the report as a single task.
4. Functionality of the current system can be increased by eliminating the use of ad hoc tools to improve information flow and, potentially, increase automation. For example, programming directly into the system application programming interface would allow required information to be extracted in more automated fashion, thereby eliminating the need to release the information to two different systems (decreasing the effort by half).
5. The contractor can consider upgrading to a fully-integrated, commercially-available project management software with features capable of streamlining the reporting process. No longer required to perform NW or PW activities, the project team will have more time to dedicate to VA tasks.
6. The contractor should prepare a list of reporting requirements and work closely with the client to mutually understand what is needed and what adds value at the early stages of the project. If completed before signing the contract, the contractor can attempt to address redundant and excessive reporting tasks with the client to potentially eliminate them as requirements.
7. The contractor can consider working with more sophisticated subtrades that are capable of providing information in a digital format (i.e., not paper-based) in future projects to reduce the amount of manual data entry.

8. Creating a lessons learned portal for project reporting to obtain assistance for future endeavors and to provide training for less experienced team members.

Using a team-based approach, the proposed process improvement solutions were discussed, and constraints limiting each proposed improvement were analyzed. Based on the feasibility of the proposed solutions, experimental scenarios for the future-state map were developed. For example, while the project team understood the value of upgrading and integrating their information systems (Improvement 5), senior management was, at present, hesitant to invest in the commercial software required. Due to the complexity, number of procedural changes required to implement certain improvements, and difficulties in accurately assessing the expected impact, Improvements 5 through 8 were excluded from the experimental scenarios proposed for evaluation. For each scenario, the impact of the proposed improvements on the design of the overall process, task durations, number of process steps, and the likelihood of rework was estimated. A summary of experimental scenarios is detailed in Table 3.7.

Table 3.7: Summary of experimental scenarios

Scenario	**Improvements Applied**								**Selected for Evaluation?**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	
1	x								Yes
2	x	x							Yes
3	x	x	x						Yes
4	x	x	x	x					Yes
5					x	x			No
6	x	x			x	x			No
7	x	x			x	x	x		No
8	x	x			x	x	x	x	No

3.5.3 Performance Prediction and Implementation

3.5.3.1 Simulation-Based Experiments

The current-state simulation model was adapted, as described in the Framework Application section, for each experimental scenario (Table 3.7), and the potential benefits were quantified using the performance metrics summarized in Table 3.3.

3.5.3.2 Future Process Development

Scenario 4, which incorporated Improvements 1-4 simultaneously (impact of scenario 4 on the process is detailed in Appendix C), had the lowest total duration and VA time of all experimented scenarios. Scenario 4 consisted of 28 process steps. The performance metric results were compared with the current-state (i.e., baseline) model as summarized in Table 3.8. The proposed improvements reduced the total duration by 31.4%, the VA time by 12.5%, and the NW tasks by 7.1%. The largest reduction in time was in the PW category, where a 94% reduction in time compared to the baseline result was observed. This is likely attributed to a reduction in the number of review rounds and to a reduction in the likelihood of rework (i.e., revisions). The NW category was associated with the lowest reduction in time (as a percentage), demonstrating that many of the activities that do not add value are necessary and unavoidable.

Table 3.8: Comparison of performance metric results from current state (i.e., baseline) versus future state (i.e., improved) simulation models

Performance Metric[1]	**Simulation Statistics (Mean Value)**		
	Baseline	**Improved**	**Reduction (%)**
Time Total	3912.8	2685.0	31.4
Time VA	398.7	348.0	12.5
Time NW	2443.9	2270.0	7.1
Time PW	1070.2	64.0	94.0
Rework Total	174.2	65.0	62.7
Rework Client	65.8	48.1	27.0
Rework Internal	108.4	48.0	55.6

[1]VA: value-adding, NW: necessary waste, PW: pure waste

The performance of each resource (i.e., team member) as a ratio of hours spent in the experimental model to the total duration of the baseline model was categorized by VA, NW, and PW in Figure 3.4. While the project coordinator spent the most amount of time in the baseline model, (1,726 hours or 44% of the total time; Figure 3.3), the time spent performing tasks in the experimental model was reduced by 48% to 887 hours (Figure 3.4). Instead, the largest amount of time spent in the experimental model was the project accountant, who was estimated to spend 1,302 hours on administrative tasks. Notably, the time spent by the project accountant was only reduced by 3.3% (Figure 3.4) in the experimental model. The greatest reduction in PW time was ascribed to the project controller (98%) and project coordinator (92%; Figure 3.4).

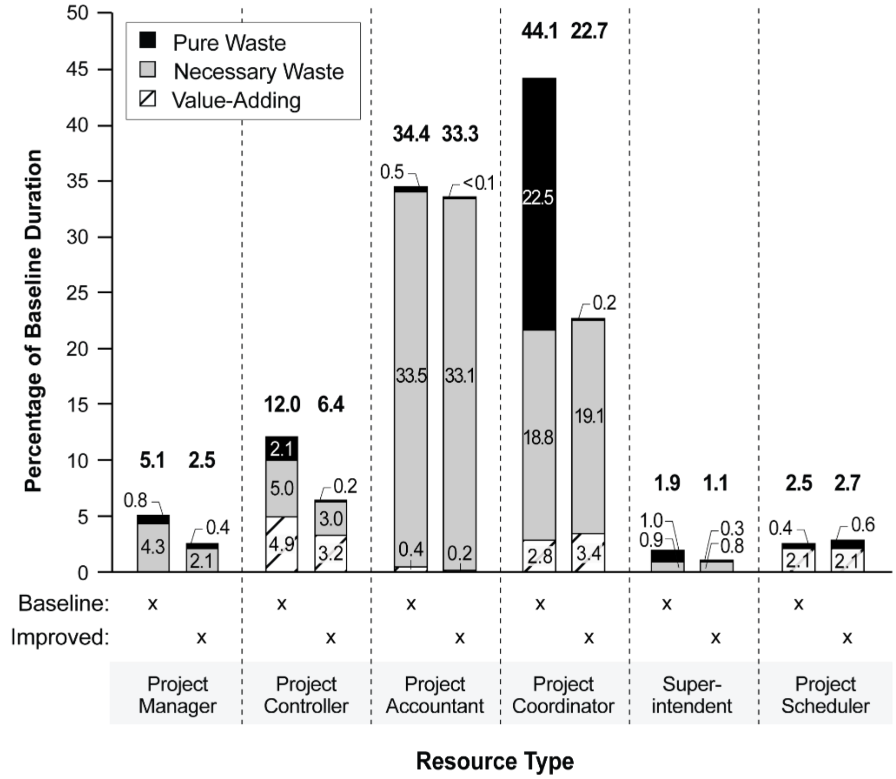


Figure 3.4: Mean value of the time spent on pure waste (black bars), necessary waste (grey bars), and value-adding (hatched bars) tasks by resource type for the baseline (i.e., current state) and improved (i.e., future state) value stream maps as a percentage of the total duration of the baseline (i.e., current state) administrative process

3.5.3.3 Selection and Implementation

Ideally, process improvement requires a continuous effort until the ideal state—where only VA activities exist—is achieved. However, in this case study, a more realistic representation of progression for short-term improvement goals that could provide fast results was explored. As mentioned previously, it was not feasible to adopt the changes required to implement Improvements 5 through 8 in the current round of improvement. Based on the results and relative ease and speed in which Improvements 1 through 4 could be implemented, Scenario 4 was chosen as the future-state map. Following implementation and testing, the proposed framework will be re-applied to evaluate the actual impact of the implemented changes and to identify any new areas of improvements.

3.6 FRAMEWORK EVALUATION

Desired objectives of this research were to (1) improve administrative efficiency in construction, (2) establish performance indicators in construction administration, and (3) develop a method capable of quantifying proposed improvements. This study develop a framework capable of stochastically modeling administrative processes and using performance metrics to quantitatively assess the performance of current administrative processes, identify potential improvements to current processes, and quantitatively predict the impact of proposed improvements on future project performance.

The developed framework was applied to a real case study and was found capable of achieving the desired objectives by successfully (1) identifying areas of improvement (Section 5.2.2) that resulted in enhanced simulated administrative efficiency (Figure 3.4), (2) establishing and applying generic performance indicators capable of assessing current and predicting future performance (Table 3.3), and (3) quantifying current (Figure 3.2-Figure 3.3, Table 3.6) administrative performance and the impact of the proposed improvements on future performance (Figure 3.4). Subject matter experts involved in the case study project confirmed the validity of the framework, indicating that the results output by the framework were consistent with what had occurred (i.e., current state) or what is expected to occur (i.e., future state) in practice. Notably, the proposed framework could be enhanced through the development of a user-friendly tool that would facilitate its application in practice.

3.7 DISCUSSION

As a newly-emerging field of research, the application of Lean principles to improve administrative processes in construction remains relatively unexplored. Of the few (i.e., less than 10; Table 3.1) studies that have explored this topic, the majority are case studies that provide specific solutions applicable only to a particular organization or distinctive administrative process. A reproducible, well-defined, and more holistic approach for the application of Lean principles to administrative processes in construction had not been developed, and as importantly, a method capable of quantifying the expected improvements proposed in these previous or future studies had not been investigated.

This study has addressed current research gaps to develop a reproducible approach for identifying areas of improvement in construction administration and quantitatively predicting their impact on project performance. Integrating Lean principles and simulation techniques, the proposed framework was able to considerably facilitate the streamlining of administrative processes in a case study of a real construction project to improve duration, efficiency, and, in turn, cost effectiveness. Application of the framework reduced the number of steps involved in the administrative process by 26%, resulting in a 31% reduction in the simulated total duration and a 94% reduction in the simulated PW time compared to the current process.

3.7.1 Theoretical and Practical Contributions

The framework creates applicable solutions to practical administrative problems in construction with theoretical relevance. The developed framework capitalizes on DES to enhance the limitations of VSM, providing a means to quantitatively assess the performance of current administrative processes, identify potential improvements to current processes, and quantitatively predict the impact of proposed improvements on future project performance. As the first study to couple DES with VSM to quantify administrative processes in construction, the proposed framework is able to stochastically model construction administrative processes and quantitatively assess current and future performance—two features that are not possible using existing methods.

Simulation, experimentation, and evaluation of the future behavior of newly-improved processes can support managerial decision-making to enhance resource planning and allocation, reduce inefficiency and redundancy, and improve overall project outcomes. The framework also enables low-risk experiments to be performed to confirm the future-state design prior to implementation with minimal impact to the company. Also, results of both the current and future-state maps can provide a baseline for other studies that are interested in applying the framework to other administrative activities.

In addition to the direct benefits garnered from the application of the framework, this research provides an opportunity to direct attention on administrative processes in construction. Construction companies have long relied on traditional approaches focused on improving production processes, making it difficult to gain the managerial commitments required to implement Lean office principles. This reluctance has been attributed to an inability to quantitatively predict the magnitude of the benefits and gains that can be achieved by implementing Lean office principles. By enabling the quantification of potential impacts, the framework can provide the evidence necessary to convince management to support the implementation of Lean principles in construction administration. Indeed, the results of this case study alone highlight the importance of the application of Lean office in construction and how Lean construction concepts and simulation modeling can be used to improve visibility and performance of administrative processes.

3.7.2 Limitations and Future Work

The findings of this study should be considered in the light of the following limitations. First, although designed to be generic, performance metrics may need to be adjusted for specific administrative processes. For example, performance metrics related to rework may not be present in some administrative processes or may be limited to internal rework requests (where rework is not requested by client). In addition, proposed improvements should be established for each new project. While previous improvements can be used as a guide, improvements may not be applicable to all projects or organizations.

Second, the validity of the quantitative results obtained by the proposed framework relies on the quality of input data. Given that companies often do not collect performance data of administrative processes, comparison with real outcomes may not be possible. In these instances, as was the case for the current case study, face validation by subject matter experts should be undertaken. Future work should explore the development of simple, quick methods for collecting these data in practice to help manage administrative processes. Decisions-makers often lack such important information and, consequently, uncertainty is present during decision-making. Furthermore, the reluctance to implement many Lean improvements arises because it is difficult for companies to predict the magnitude of the benefits and gains that can be achieved by implementing Lean principles.

Third, the goal of this study was to improve administrative processes to enhance the production variable. As such, this work focused on production metrics, such as time, and process complexity metrics, including VA, NW, and PW times, process steps, and VA process steps. The production variable has the potential to influence the performance of other variables, such as cost and quality. Future work should include the development of approaches capable of quantitatively assessing such variables and evaluating the impact of process improvements on the cost and quality of the process.

3.8 Acknowledgments

This research work is funded by a Collaborative Research and Development Grant (CRDPJ 492657) from the Natural Science and Engineering Research Council of Canada. The authors would like to thank Graham Industrial Services LP for their support and for providing contractual documents and report preparation-associated data. The authors also would like to acknowledge Catherine Pretzlaw for her assistance with manuscript editing and composition.

Chapter 4

Social network analysis of contract-related documentation processes for communication assessment

4.1 Introduction

In a construction project, a network of organizations work together and enter into various communication arrangements to create value and achieve project goals (Dietrich *et al.*, 2010). These organizations are interdependent and influence each other. They need to exchange large amounts of information about activities, processes, and decisions used to deliver the project. In project-based organizations, communication plays an important role in project success. Although the impact of communication on projects has received attention in academic research (Smit *et al.*, 2017), there has been limited research on the effects of communication and coordination across different organizational levels for contract-related documentation and reporting. Communication and coordination can be viewed as valuable resources or intellectual assets. The communication performance of project participants is an important factor for efficient project reporting and must be assessed periodically during the execution of the project to identify bottlenecks, enhance performance, and improve the reporting process.

Change order reports, as one type of construction report, require effective communication between diverse project participants to control delays and costs. Throughout the lifecycle of a construction project, iterative cycles of changes are common; these changes can cause uncertainty and complexity in construction management (Hao *et al.*, 2008; Lee *et al.*, 2006; Motawa *et al.*, 2007), resulting in significant cost overruns and delays. A change can refer to any variation or a modification to existing conditions, assumptions, or requirements in construction work (Sun *et al.*, 2004). Changes usually lead to issuing change orders to deal with variations in the scope of work, such as material or design. The implementation of change orders can cost approximately 5.1–7.6% of the total project (Cox *et al.*, 1999). A review of change orders in construction has shown that changes can range between 10–15% of the contract value and cause 10–20% losses in productivity. A 40% increase in project duration was observed in one case study (Desai *et al.*, 2015). The full cost of changes increases nonlinearly with the cumulative size of all changes (Cooper and Reichelt, 2007), and a change control system is required to control the overall management of change orders (Ibbs *et al.*, 2001). While several studies have developed change management systems for construction projects (Ibbs *et al.*, 2001; Karim and Adeli, 1999; Lee and Pena-Mora, 2007), the importance of project participants in the change management process has not been widely studied. Construction projects require collective effort, communication, and coordination among project participants, especially in the change order process of a project (Butt *et al.*, 2016).

This research emphasizes the impact of communication and information exchange between participants to enhance efficiency. Project participants not only exchange project information as formally determined, they also continuously exchange knowledge and insights to enhance collective project performance. The network of this information exchange can be quite complex and impossible to analyze manually. Therefore, there is a need for big linked data analytics among communication networks to effectively support project success. This study explores change order communications; describes how individuals are engaged in and impact the communication network; and emphasizes the significance of social processes, patterns, and practices in process-efficiency enhancement. Social network analysis (SNA) has been the main approach adopted within multi-organizational networks to identify and analyze participant relationships, roles, and overall network structure (Zhang and Ashuri, 2018). SNA is a powerful tool to study complex systems. Here, SNA is employed to measure and analyze the communication of change order information. A case study project provided by a construction company in Alberta, Canada, is used to examine the communication performance in a multi-organizational working system. The findings contribute to an understanding of how communication impacts the change order process and, in turn, the project.

The existing literature on change management and communication impacts is reviewed. A systematic approach to identify the social networks embedded in the change management process is then presented. The concept of social network is introduced and network modeling at different organizational levels is explained. Finally, a case study provided by a construction company is discussed, followed by a discussion of limitations, conclusions, and directions for future research.

4.2 Research Background

4.2.1 Change Orders

Change orders are the most common factor for cost and schedule overruns in construction projects (Shrestha and Zeleke, 2018), and they often impact the project quality, time, and cost. Several studies have examined the effects of change orders on different aspects of construction, including labor productivity (Kermanshachi et al., 2018) or cost and schedule overruns (Serag et al., 2010).

These impacts can be attributed to multiple causes, including variations in project scope, lack of project communication, poor site-management, improper planning of material quantities, or vendor changes. Change orders can strain the relationships of the owners, contractors, subcontractors, and other organizations involved in the construction process. Lee et al. (Lee *et al.*, 2005) proposed a dynamic planning and control system to evaluate the negative impacts of changes and other conflicts on the construction project performance. Zhao et al. (Zhao *et al.*, 2010) developed a methodology to predict changes due to the responsible factors from information flow. Moselhi et al. (Moselhi *et al.*, 2005) and Ibbs (Ibbs, 2013) focused more on the cumulative impacts of changes, and they showed that disruption due to changes overshadows productivity. Many researchers have stated that lack of communication leads to change orders and rework (Alnuaimi *et al.*, 2009; Bröchner and Badenfelt, 2011; Safapour and Kermanshachi, 2018; Sun and Meng, 2009). However, past research lacks quantitative analyses of impact that project participants have on the change-management process.

4.2.2 Communication in Change Order Reporting Process

Project participants are often geographically dispersed while they exchange a massive amount of information to enhance project success (Butt et al., 2016). Communication is of utmost importance to coordinate goals in construction projects. While much research has been conducted to study the role and benefits of good communication on the overall success of a project (Dainty et al., 2007), little work has been devoted to studying the effects of communication in more detailed aspects or processes, such as those involved in the change-order reporting process. Recent work in the construction and project management literature has shed light on the importance of good communication to achieve success in projects (Dossick et al., 2014; Henderson et al., 2016; Manata et al., 2018; Smit et al., 2017; Turner and Müller, 2004). Ineffective communication for change order by one party can cause disputes and delays of work of other parties in the project. Padalkar and Gopinath (Padalkar and Gopinath, 2016) conducted an extensive literature review and found that although there was an increase in project communication research between 2011 and 2015, communication is a minimally represented knowledge area. Smit et al. (Smit et al., 2017) investigated communication preferences of project participants to evaluate how modern communication media impact project participants and outcomes. Butt et al. (Butt et al., 2016) presented a qualitative study for understanding the effects of communication between project stakeholders on the change management process in two construction projects. They concluded that effective communication created clear change management processes that found innovative solutions for problems. Charoenngam et al. (Charoenngam et al., 2003) developed a web-based tool to establish a good communication framework between project stakeholders for effective management of change orders. At this time, there is no suitably developed framework for analyzing the communication performance of project participants in a change order in the construction industry. Three centrality measures are used in this study to compare and evaluate the performance of each participant in the network – i.e. how involved a participant is in the communication (degree centrality), how a participant controls the flow of information (betweenness centrality), and how active a participant is in exchanging information (closeness centrality). This paper deals with the assessment of communication to fill this gap in the construction domain.

4.2.3 Social Network Analysis

The characteristics of communication networks involved in the change management process can be studied using social network analysis (SNA) (Zhang and Ashuri, 2018). The concept of social network analysis was first introduced by Moreno (Moreno, 1960) to study social interactions, and has recently been utilized in the fields of engineering and construction (Park et al., 2010). Social network analyses broadly identify social structure interactions. Graphs or sociograms are created with nodes representing the parties in a network and links between the nodes representing the relations between the parties. Social network analysis emphasizes the relational measures among the parties represented in a graph or sociogram. Many researchers use SNA to identify and analyze structural properties of various relationships in the construction management domain (Chinowsky and Taylor, 2012; Dogan et al., 2013). SNA can also be used to compare project participant performance as predefined and shaped by contractual agreements and communication links. This approach appeals to researchers in the construction domain because of its capability to investigate various relationships among project participants and organizations.

4.3 Research Framework

Change management data contain all of the change orders and activities requested or ordered by project participants. This section focuses on the application of SNA to analyze communication among project participants in the change management process. Change management data are used as a means to discover social networks in these communicative processes. The main challenge in using change management data is that tremendous volumes of change orders are stored in unstructured text formats with both noise and outliers, making them difficult to process and analyze. Figure 4.1 illustrates the framework for discovering social networks from change management data. The proposed framework is detailed in the following sections and is composed of three main steps: (i) social network data mining, (ii) social network modeling, and (iii) social network analysis. The change management data are used as a data source in the first step of the network mining. The next step is to extract the required data for social network analysis, which is the first step of data generation.

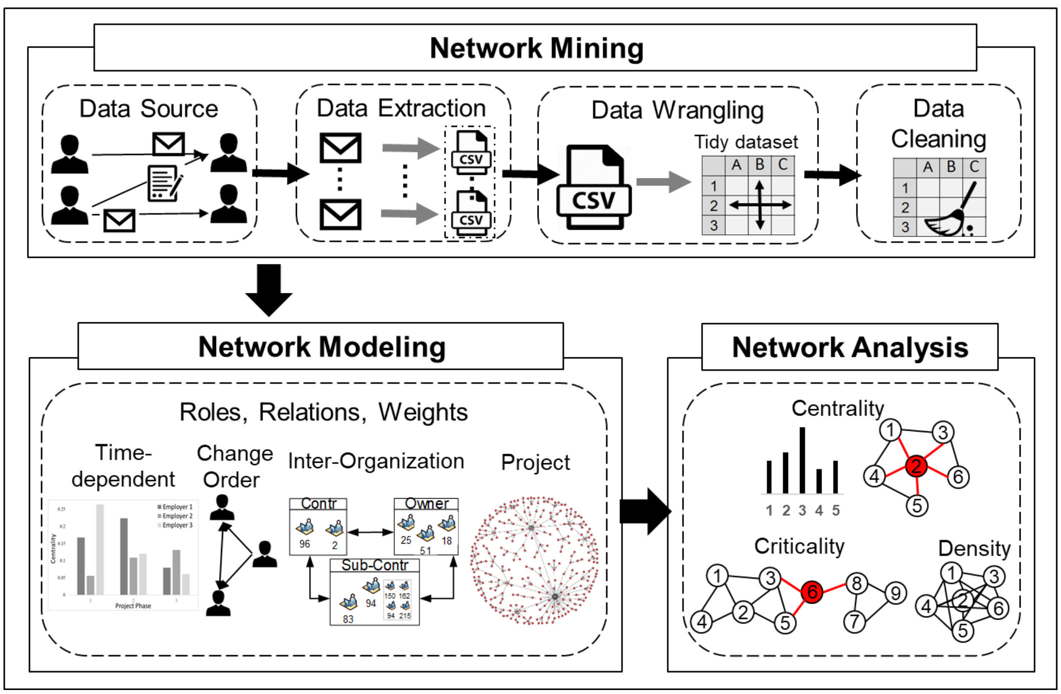


Figure 4.1: Research framework to analyze communication

4.3.1 Network Mining

The data mining procedure first consists of data extraction, wrangling, and cleaning to obtain change order communications from several participants over the course of a project in a comma separated values (CSV) format. The challenge is to mine data for a directed network consisting of n participants, which requires identification of the properties of $n(n-1)$ pairs of participants. Any change order requires multiple participants to work together to address the requirements. Change orders often exist simultaneously, leading to a substantial number of communication links. To streamline change order communications from different project participants, each link is transferred into CSV format.

In the data extraction step, all the information related to the participants in the change order communication network is extracted from change management data. Information about participants is saved using the participant's name and organization (or a unique ID number) in the extracted CSV file, and each line in the dataset represents an interaction between multiple project participants. Several other items such as the change order number/description and timestamp are pulled out and stored in the CSV file, as well. For example, Participant #2 sends a letter to Participants #25, #92 and #94 regarding Change Order #2 in June 2013. This observation in the extraction step of data mining is stored in one row with multiple columns; each column represents a feature related to the observation. The information in this row is relevant to three different observations and should be reflected in three communication links in the network in the data wrangling step. In data wrangling, the dataset is reshaped and combined into compatible and interpretable formats. All CSV files are combined into one file containing all related information regarding change orders and participants. In the final file, each line represents an interaction between two participants. In the data cleaning step, data points that are not true (noise and outliers) for the communication application are removed in the data cleaning step. For example, a communication between a participant and themselves is an outlier in the dataset, and null or missing values (i.e. blank email messages) are noise. Table 4.1 illustrates a sample of the extracted CSV file after the data-mining process.

Table 4.1: Sample of the extracted CSV file

Sender ID/ Organization	**Receiver ID/ Organization**	**Timestamp**	**Subject**
92/ Contractor	25/ Owner	3/7/2013 11:49:38	“COR 03 Pending action”
54/ Contractor	144/ Subcontractor2	14/12/2013 6:42:48	“COR 10 Lake travel allowance”
94/ Subcontractor1	92/ Contractor	11/1/2014 13:38:44	“Executed COR 19”
144/ Subcontractor2	54/ Contractor	15/12/2014 6:27:52	“COR 42 Snow removal”
137/ Subcontractor2	92/ Contractor	9/3/2015 10:36:30	“COR 114 Tanks handrail”
25/ Owner	92/ Contractor	8/4/2015 9:43:17	“COR 152 Approved”

4.3.2 Network Modeling

Social network modeling develops a sociogram representing relations among multiple participants in a network. To create a social network, three main components need to be determined from the change management data: participants, relations, and weights. Participants of the network can initially be identified in the traces left behind in the change management data. Relation weight is measured by the total number of links sent by a participant and received by another participant in the network. There is no standard or metric to construct social networks as they are tailored to the application or the objective of the study. SNA is conducted to measure the structural characteristics of the communication network at the following levels:

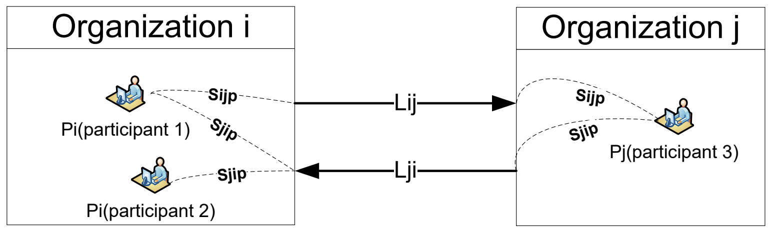
4.3.2.1 Project Level

At the project level, SNA analyzes and measures characteristics of a network to enhance the understanding of participants' communication from the perspective of the whole project. Moreover, SNA is used to evaluate the position of each node in the project by calculating centrality measures as will be described in the Network Analysis section below. Three centrality measures (degree, betweenness, and closeness) are typically considered in SNA.

4.3.2.2 Inter-Organization Level

In order to accomplish a multi-organizational project, organizations must be willing to communicate, coordinate, and share their knowledge. In construction, a network of organizations works together to achieve the goals of the project and to tackle the change orders. Organizations are usually geographically separated but still make interdependent decisions. Effective communication among involved organizations is, therefore, vital for smooth change-order execution, to maintain compatibility, and to achieve project objectives. At the inter-organization level, only links between organizations are needed to evaluate a social network. Social network analysis can be used to depict both formal and informal organizational links. In this research framework, a communication network between different organizations is built based on the cleaned data summarized in Table 4.1. To calculate the strength/contribution of a participant in the inter-organization level, the authors created a percentage calculation. First, let L_{ij} represent total number of links sent by organization i to organization j. A communication matrix is built to represent the weighted relation of links among the main organizations, as shown in Figure

4.2**Error! Reference source not found.**. For example, the weight of the link from Organization 2 to Organization 3 is calculated to be L_{23}.



$$L_{ij} = L_{ijp(participant\ 1)} + L_{ijp(participant\ 2)}$$

$$L_{ji} = L_{jip(participant\ 3)}$$

Figure 4.2: Inter-organization communication

Next, let p be the ID of a participant within organizations i or j. Let p_i be the ID number of a participant within organization i and p_j be the ID number of a participant who works for organization j. L_{ijp} is the total number of links sent if p belongs to organization i and received if p belongs to organization j. The calculation of the strength/contribution of a participant is given by Equation (4.1)

$$S_{ijp} = L_{ijp}/L_{ij}, \ \ if\ L_{ij} > 0\ \&\ i \neq j \tag{4.1}$$

where S_{ijP} serves as an indicator to measure the strength or contribution of participant p $(p = 1, 2, \dots, P)$ on communication links L_{ij} $(i, j = 1, 2, \dots, N)$.

At this level, the strength, betweenness, and closeness centrality measures refer to organizations, rather than to individual participants. Here, strength centrality measures the extent to which an organization is connected to adjacent organizations. An organization's in-strength and out-strength centralities represent the degree to which the organization is a receiver or sender, respectively, of information from or to the organization's neighbours. Betweenness centrality measures the extent to which an organization controls the flow of information between different organizations. Critically, these interstitial organizations keep the network together (Wasserman and Faust, 1994). Closeness centrality measures the distances of an organization to every other organization in the network, reflecting its dependency/independency. An organization with high closeness centrality is highly dependent, making it difficult to act independently without others knowing. A centrality index is defined for each company using the average of the three centrality measures. The centrality index of a company—and the number and direction of the links exchanged between the participants—could be measured easily using a programming language, such as R (Team, 2013). The centrality measures of an organization indicate the position, intensity of power, and influence of this organization in the network.

4.3.2.3 Change Order Level

Typically, when a change order is identified in a project, the responsible organization will try to evaluate and quantify its impacts. Depending on its complexity, the schedule and cost impact of a change order can be estimated; however, the cumulative impacts of the change order or its impact on downstream activities is not as straightforward to calculate. Participants themselves may impact the duration of the change order, based on their individual proficiencies. SNA at the change-order level evaluates the position of project participants involved in processing a specific change order. To identify the network for a specific change order, Figure 4.1 needs to be filtered based on the subject of a change order. The filtration process results in the identification of participants, their organizations, and timestamps of the communication links. The involved participants and the duration of their actions is then identified and evaluated in each step during processing. By analyzing the network at the change order level, different stages to a change order are revealed. The results also provide insight into delays due to the processing time between liable participants. In each step during processing, the involved participants and the duration for their actions is identified and evaluated. The results provide insight into bottlenecks in the process, creating a decision-support tool for project managers that will enable them to improve their change management system.

4.3.2.4 Time-Dependent Level

At this level, SNA is applied to evaluate the network of the change-management process at different project phases. Participant roles may change during the construction phases of the project. Some participants, such as a cost engineering managers, play more important roles in project initiation and project closeout. A procurement specialist is usually more involved at the peak of construction, which might require more change orders related to procuring different materials. Here, SNA shows the centrality of the most important participants over time. The results provide insight into participant performance during construction.

4.3.3 Network Analysis

This section clarifies the network analysis calculations and metrics that are conducted at different levels of the change order. These metrics include network descriptions, degree centrality, betweenness centrality, and closeness centrality.

4.3.3.1 Network Characteristics

SNA provides an effective tool to measure network size, density, average degree, etc. The elements of $N = \{1,2,\dots,n\}$ are the nodes or participants of the network, while the elements of $L = \{l_1, l_2, \dots, l_K\}$ are its links or edges. A directed network consists of two sets *N* and L, such that $N \neq \emptyset$ and L is a set of pairs of *N*. Here, a node is referred to by its order, *I*, in the set, *N*. In a directed network, each link is defined by a couple of nodes, i and *j*, and is denoted as l_{ij}. The order of the two nodes is important: l_{ij} stands for a link from *i* to *j*, and $l_{ij} \neq l_{ji}$, with $l_{ij} = 0$ representing the absence of an edge from i to *j*. Network size is the total number of nodes or ties. Network density is the proportion of existing connected ties over all the possible connections (see Equation (4.2)). Networks with high density are highly connected, and information or resources can quickly move across the network. Network density is a representative of the cohesion of the entire network and can be used to provide more insight into how connected participants from different organizations are at the project level. Average degree measures the average number of neighbors per node (see Equation (4.3)). The average degree is closely related to the density.

$$D = \frac{L}{n(n-1)} \tag{4.2}$$

$$AD = \frac{L}{n} \tag{4.3}$$

Here, *D* is the density of a directed network, *L* is the number of existing links, n is the total number of nodes/participants in the network; *AD* is the average degree. The diameter is the longest path of all the calculated shortest paths in the network. In other words, diameter is the largest distance between the farthest nodes in the network. In SNA, distance is calculated by the number of links in the shortest possible paths from one node to another. The mean distance provides a measure of communication efficiency for an entire network by averaging the shortest possible path between all nodes. The measure of reciprocity defines the proportion of mutual connections, in a directed graph. In other words, it is an index to measure the tendency of participants to reciprocate. It is most commonly defined as the probability that the opposite counterpart of a directed link is also included in the network. Social network analysis facilitates comparison between project participants' actions within their communication channels. Network centrality measures describe the intensity of power, prominence, and influence of a network participant.

4.3.3.2 Degree Centrality

Degree centrality is the number of adjacent edges or ties a node has as a participant representative. It is an indicator of how connected a network participant is to other participants. A higher degree centrality indicates higher interaction, more influence, and stronger involvement in the network. Degree centrality is calculated using Equation (4.4)

$$DC_i = \frac{\sum_{j=1}^{N}(L_{ij} + L_{ji})}{2(N-1)}, \quad 0 \leq DC_i \leq 1 \tag{4.4}$$

where DC_i is the degree centrality of the *i*th node in network, L_{ij} is the number of links that a node *j* receives from a node *i*, *N* is the total number of the nodes in the network. The weight or importance of an arc is not reflected in the degree centrality. For a weighted network, strength centrality is a better representation of each node connection or links. A weight of the relationships can be used to represent the strength of the relation between participants in a communicative environment. Strength is calculated by summation of the weight of adjacent edges. Therefore, it counts for both the number of adjacent edges and their weights. If the strength of the relation is required, the weight of each relation should be taken into account. DC_i measures the node degree centrality by the sum of the weights.

4.3.3.3 Betweenness Centrality

The betweenness centrality of a node is the number of times a node acts as a bridge along the shortest path between two other nodes (see Equation (4.5)). In other words, betweenness centrality measures the extent to which a participant is located in the shortest path between two other participants in the network; potentially, this participant controls this flow through the network. Participants with a high betweenness centrality occupy critical network positions; poor performance by participants at these critical points can harm the network. A higher betweenness centrality may indicate a participant who takes on an informal leadership position in the network and may encourage a participant to contribute more to solutions in response to the problems encountered in the project. In Equation (4.5), BC_i is the betweenness centrality of the ith node in network, $\sigma_i(s,t)$ is the number of the shortest paths that pass through node i, and $\sigma(s,t)$ is the number of shortest paths between all nodes.

$$BC_i = \sum_{s,t:s\neq t\neq i} \frac{\sigma_i(s,t)}{\sigma(s,t)} \tag{4.5}$$

4.3.3.4 Closeness Centrality

Closeness centrality reflects the extent to which the network is concentrated around one participant. The closeness centrality of a node is the average length of the shortest path between one node and all other nodes in the network. The more central the node, the closer it is to all the other nodes. A node/participant with a high closeness centrality can be very active and quick in exchanging information with other participants in the network. Calculation of the closeness centrality is given by Equation (4.6)

$$CC_i = \frac{n-1}{\sum_{k \in N} d(i,k)} \tag{4.6}$$

where CC_i is the closeness centrality of the ith node in network, and $d(i,k)$ is the length of the shortest path (geodesic distance) between nodes i and k. In problem solving that relies on communication links, efficient solutions occur when a participant has the shortest communication paths to the other participants. Communicating with a participant with high closeness centrality can be accomplished in an easier, more direct, and more efficient manner.

4.4 Case Study

A large dataset of change orders for an oil and gas project was provided by a construction company located in Alberta, Canada, and used to demonstrate the feasibility and applicability of the developed approach in this research. The construction company played contractor role in a design-bid-build project to build an industrial plant. The major stakeholders in this project were the owner, main contractor, and subcontractors 1–3. The main contractor was responsible for delivering the project on time and budget to the owner. This project was a lump-sum contract with an original contract value of approximately 700 million dollars (CAD). The contract price was increased by 150 million dollars due to the change orders.

To create the network, change orders were read automatically to capture the records of communications. After the data cleaning step, a total of 3,402 communication links were generated among different project participants dispersed across different organizations and stored in a CSV file. Each line in the file corresponded to information about a single sender/receiver set, a timestamp, and a change order description, as shown above in Table 4.1.

Two main components are required to create a social network: participants and their relationships. Project participants were identified by extracting unique IDs using the CSV file and were then stored in a separate CSV file, the node file. There were 412 nodes (participants) in our node file. Each node was a representation of a unique participant in the change-order network who worked for one of the organizations: contractor, owner, or subcontractor. Each was marked with a unique label. After each participant was labeled, the links between nodes were saved in an arc file, which contained the 3,402 links between labeled nodes.

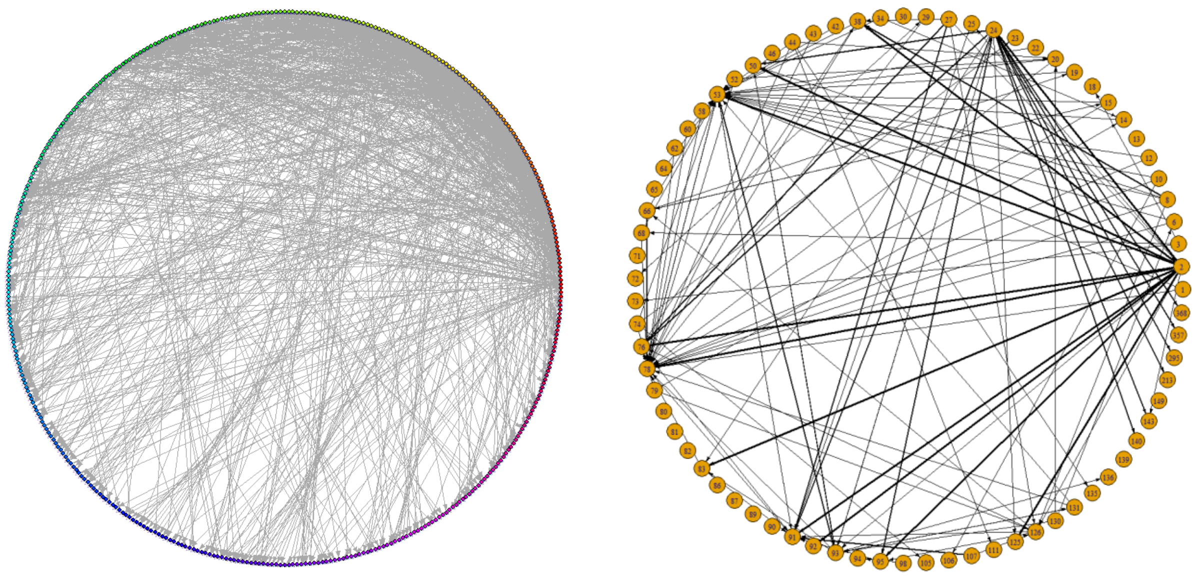
4.4.1 Results

In this section, the node and arc files are utilized to visualize and interpret the social network of communication. These facilitate better understanding of communication characteristics between participants at numerous levels.

4.4.1.1 Project Level

At the project level, the cleaned dataset contains information about 412 participants. Figure 4.3 shows the social network built based on the relations among project participants. The communication network description is a directed network with 412 nodes and 3,402 arcs. Figure 4.3(a) represents all participants involved in the entire project and all of the communications between them. In Figure 4.3(b), only nodes with strength centralities greater than 20 are shown.

The widths of the links correspond to the weight/frequency of communication. Table 4.2 summarizes the communication measured at the project-level.



(a) All nodes and relations (b) Nodes with strength>20

Figure 4.3: Social network at project level

Table 4.2: Characteristics of the network at project level

Symbol	Description	Value
N	Number of nodes	412
E	Number of edges	3402
D	Density of network	0.008
R	Reciprocity of network	0.142
DM	Diameter of the network	23
MD	Mean distance of the network	4.14
AD	Average degree of the network	6.49
AS	Average strength of the network	16.51

As provided in Equation (4.2), network density falls within the range of 0–1. At a value of 1.00

(highest value), all participants are connected to each other; the network has total connectivity. At a value of 0 (lowest value), none of the participants are connected; a network is absent and all participants are isolated. Here, the network density is 0.008, which shows that in this network only 0.8% of all possible links among project participants are present, suggesting a low level of network cohesion. The total number of nodes/participants was 412, which is very high and indicates that there are many unrealized, potential relationships in the network. The average degree of the network is 6.49, indicating that each participant among the 412 participants in the network is connected to six to seven other participants on average. For a weighted network average, strength is a better measure than average degree because strength accounts for both the number of links and their frequency. In this case, the average strength of the network is 16.51 – almost 2.5 times the average degree – which could be a result of the intention/preference of participants to make communication with the previously interrelated participants in the network. Considering the close relationship between average degree and density, the results indicate that the structure of a network can be affected by a high concentration of links in a few participants, even if the other participants have few connections as shown in Figure 4.3. The value of the network diameter is 23, representing the distance between the two farthest nodes in the network. The mean distance of the network is 4.14, meaning one node could traverse 4.14 nodes to touch another node. The reciprocity of the interactive network is 0.142 which shows the network is poorly reciprocal.

As mentioned earlier, centrality measures are used to evaluate the embeddedness of nodes and their positions in the discovered social network. Here, the results show that Participant #2 has the highest degree of centrality in the project level, indicating their high activity and involvement in the network. Participant #2 is the procurement specialist working for a contractor who is the sender and receiver of a large number of communication links. Participant #66 has the highest betweenness centrality and is able to control the communication flow easily. Participant #66 is a senior project manager for the contractor whose higher betweenness value indicates a leadership position in the network. It is highly effective for Participant #66 to contribute to solutions in response to the change order problems encountered in the project. Participant #174 has the highest closeness centrality and is very active in communicating information to other participants.

When the project is studied at the change order level, there are many people involved, which makes it difficult for human eyes to recognize the most important participants in the visualized network shown in Figure 4.3(a). However, the configuration of the network and centrality measures of the participants can be easily seen in a tabular format (

Table 4.2). In particular, strength centrality can be used to reveal the key participants and connections that have a significant impact on the change order communication at the project level. The results of the centrality measures, however, cannot be generalized to all projects because each project is unique; the change orders will be completely different. It is also expected that there would be a connection between the most frequent change orders and participant responsible for those types of changes due to the nature of their positions. In this case study, for example, most of the change orders are related to changes in material, which is why the procurement specialist has the highest degree centrality in the network.

4.4.1.2 Inter-Organization Level

Based on the cleaned data summarized in Table 4.1, the active participants are mostly distributed among 5 major organizations. For simplicity, a communication network has been built for the 5 main organizations (Figure 4.4). A weighted relation assesses the strength of the communications between organizations and project participants. Frequency is reflected in weighted arrows, with each arrow showing the number of communications. The contractor here communicates with the owner more than the subcontractors because the contractor is mainly responsible for delivering the project with all the changes requested by the owner and is also responsible for getting the approvals. Simultaneously, the owner has sent communications to the contractor 164 times. The sent/received links are represented as percentages of the total number of communication links to show the contributions of the participants. For example, Participant #2 contributed 85% to the communication links sent by contractor and received by owner. Therefore, Participant #2 has a stronger role in the relationship as compared to other participants who belong to the same organization. Participant #25, on the other hand, received 35 percent of the links sent by the contractor.

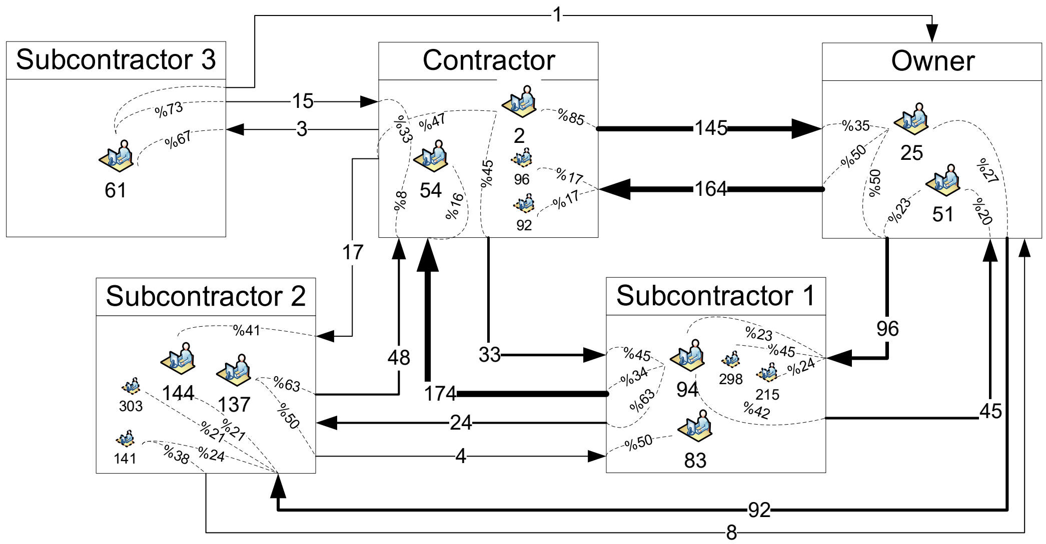


Figure 4.4: Social network at the inter-organization level

Several participants belong to tightly connected organizations, as shown in Figure 4.4. Many of the other participants are completely isolated from the inter-organization level, thus are not represented. The bridging roles that connect different organizations in the network can be identified for improving information sharing within the project. Participants #2, #25, #54, and #94 are all involved in most inter-organizational communications. These participants keep in close touch with more than one organization, acting as boundary spanners to provide an information channel among different organizations. These participants play a significant role in communication and knowledge sharing in the network. By contrast, Participants #83, #141, and #303 are seen only in one inter-organization link, indicating that they are considerably isolated. The results of centrality calculations at the inter-organization level are presented in Table 4.3. The centrality values are normalized so that they range between 0 and 1. The centrality index is used to represent the average of the three centrality measures in Table 4.3.

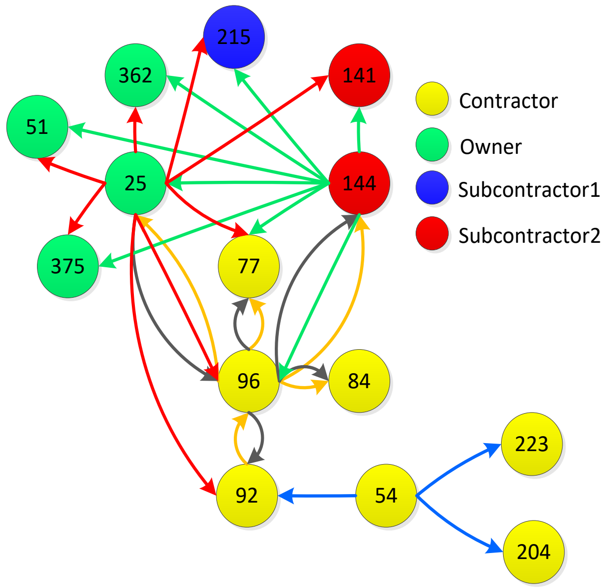
Table 4.3: Centrality measures at inter-organization level

Organizations	Normalized Centrality Measures			Centrality Index
	Degree	Betweenness	Closeness	
Owner	0.396	0.9	0.028	0.441
Contractor	0.430	1	0.278	0.569
Subcontractor 1	0.332	0.9	0.028	0.419
Subcontractor 2	0.199	0.8	0	0.333
Subcontractor 3	0.014	0.686	0	0.233

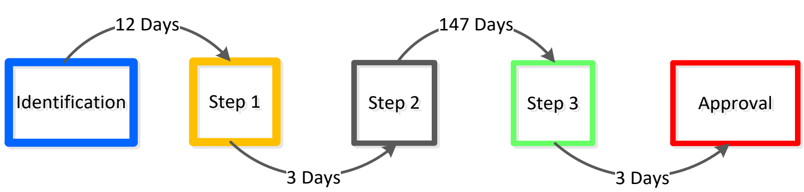
The network centrality of an organization is an indicator of its power and influence, representing how strategically an organization is connected in the network. The contractor has the highest centrality index with the highest degree, betweenness, and closeness measures. High-degree centrality for the contractor indicates high activity and involvement in the network. As shown in Figure 4.4, the contractor is connected to all the organizations in the network and is the sender and receiver of a large number of communication links. The contractor has the most central position in-betweenness measurements and is able to control the communication flow easily. In this study, a higher betweenness value of contractor indicates a leadership position in the network. The contractor here is very active in communicating information to other organization because of their high closeness centrality. Analysis of the inter-organization level shows which participant might be the bottleneck in communicating change orders between different parties in the project due to their burden. These results can help organizations to improve the communication skills related to change management process.

4.4.1.3 Change Order Level

In this section, one change order is considered as an example to show the results of SNA at the change-order level. This change order is related to the procurement of pipe. SNA is applied to evaluate the performance of the participants at the change-order level. As shown in Figure 4.5, four organizations and 14 participants are involved in these change order communications. Each participant is shown using a colored circle with a number inside it to represent the unique ID of the participant. The color of the circle provides information about the organization to which the participants belongs. In Fig 5, participants belonging to contractor are shown in yellow. In Figure 4.5(a), the colored arrows correlate with the timeline of the change order as shown in Figure 4.5(b) The change order is identified, and blue is used to show the links at identification step. It took 12 days for change order participants to analyze and send change-order information out. In Figure 4.5(b), the timeline from identification to approval of this change order is shown. There is a gap of 147 days from Step 2 to 3 in the process of the change order, which may be the result of a performance deficiency of the participants involved in Step 2. This delay may cause additional delays and costs to the project and should be used to evaluate performance.



(a) Network of change order



(b) Timeline of the change order

Figure 4.5: Social network at change order level

The participants of each organization keep in close touch with three other organizations' employees, acting as the boundary spanners to provide an information channel among different organizations. Participants #25, #96, and #144 play a significant role in communication and knowledge sharing in the network of this change order. By contrast, Participants #223 and #204 show up in only one communication link, indicating that they do not keep in close connection with the other participants. The results of centrality calculations at change order level are presented in Table 4.4. The centrality values are normalized so they range between 0 and 1. The centrality index represents the average of the three centrality measures.

Table 4.4: Centrality measures at change order level

Participant ID	**Normalized Centrality Measures**			**Centrality Index**
	Degree	**Betweenness**	**Closeness**	
25	0.212	0.051	0.366	0.209
51	0.039	0	0.228	0.089
54	0.058	0	0.276	0.111
77	0.077	0	0.236	0.104
84	0.039	0	0.23	0.089
92	0.077	0.064	0.295	0.145
96	0.231	0.138	0.347	0.239
141	0.039	0	0.228	0.089
144	0.193	0.048	0.363	0.201

The top participants in this change order ranked in degree and closeness centralities are Participants #25 (contract administrator) and #96 (project director); the top participants ranked in betweenness centrality are #25 and #92 (cost engineering managers). Participant #96 has the highest centrality index.

4.4.1.4 Time-Dependent Level

In this section, to capture the importance of participant roles at the granular level of the project timeline, the project duration is divided into seven phases from initiation to close out. In each assumed phase, SNA is applied to analyze the change-management network, and the centrality of participants is calculated. The participants with higher centralities are compared at different periods of the project. Here, the roles of five participants are evaluated at these assumed phases.

Participants #2, #25, #54, #96 and #92 are compared. The comparison results for these five participants are shown in Figure 4.6. Participant #92 is a cost engineer manager who works for the contractor whose role is more important in the first and last periods of the project compared to other participants. Participant #2, alternatively, has a more important role in the third and fourth periods of the project.

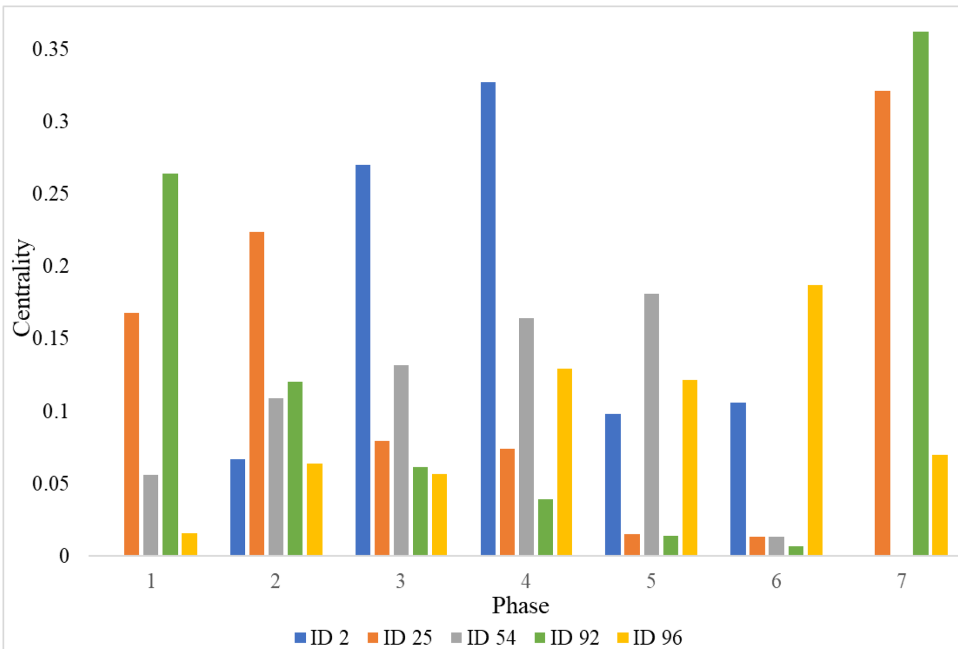


Figure 4.6: Participant centrality vs time

4.5 Discussion

The change order process is a combined effort among the owner, contractor, and subcontractors. Each participant involved in the change order network is expected to communicate the required information at the appropriate time to avoid delays. Issues and problems arise due to a lack of a clear schedule for change orders, too many arriving on a participant's desk at a given time, unclear tracking processes, and ultimately, information not being communicated in a timely manner.

SNA metrics and concepts applied in this research provide useful information about network formation, centrality, and connectedness of network. In terms of the overall network, network density is used to indicate the strength of the connections in a network. A low-density value indicated that a network focuses on individuals rather than on collaboration over a network. Such networks need to be redesigned by project managers to increase their density for a more efficient communication structure. Therefore, SNA is applied to discover underlying problems in change management communication network that can be later investigated to identify misalignments that impede project effectiveness.

Using this approach on an ongoing construction project can illustrate important communication barriers that exist in early project stages. By analyzing network communication performance, project managers can detect potential phenomena leading to communication problems and difficulties. Later, the communication network can be redesigned to involve the same participants more effectively, utilizing their skills and leading to a better change management process.

The findings in this research can help construction project managers define appropriate staffing and task allocation strategies during or after execution of a project. For example, as shown in Figure 4.4, the project manager needs to take into consideration that Participant #2 may create a bottleneck in the network. If Participant #2 is overloaded with many change orders to be processed, it may cause delays. The project manager can assign another team member to distribute the workload. Additionally, historical records of communication and actual performance provide an unprecedented opportunity for change management. Project managers can use historical records of the quantified communication in making critical staffing decisions, such as selecting the leader of the change management team or grouping the team members to work on specific change orders. Fig. 6 shows that Participant #2 has more involvement in phases 3 and 4 of the project, but their involvement in phases 1 and 7 is negligible. This can help the project manager to allocate resources over project phases.

The quantified change management communication network in this research also provides useful information to identify bottlenecks in change order process development. A change order may be delayed because the liable participant has not reviewed it; as a consequence, corresponding follow-up measures have not been taken, including reassigning the work to other available participants. For example, as shown in Figure 4.5(b), proceeding from step 2 to step 3 takes 147 days, which should be a red flag for the project manager to further investigate the reason for the delay and assess its impact on overall project timeline. Visualizing the communication features of the change management network also shows places that the change order process needs improvement. This improvement could entail replacing personnel, adding more staff, changing the configuration of the team, or assigning team members to different tasks.

The quantitative analysis of change management communication could also provide the project manager with critical information to develop targeted training programs for team members who are anticipated to run into difficulty in upcoming projects. Customized knowledge management can be developed to share the lessons learned from past communication strategies to enhance change management in future projects. Leadership training can be provided to key team members based on team decision- making in past projects and their performance in managing change orders.

4.6 Limitations and Future Work

Mining tremendous volumes of change management data that are stored in different formats to explore the communication network is a tedious task. One limitation is the need for a data adapter to harvest, wrangle, and clean the change order communication data and convert them to the required CSV format. Additionally, many organizations are involved in the change management process, which makes it difficult to develop a unified system to collect all change management data. Depending on the project delivery system and other contractual arrangements, the communication network may change, and the findings of this framework are only applicable for similar projects within a company. One other limitation is the type of communication that were used; informal phone calls or face-to-face conversations were not captured for this study.

As the case study demonstrates, identifying network participants and their links manually is not feasible. Therefore, one of the benefits of using the proposed framework is automating the network extraction and analysis. Networks from similar historical projects could be used to predict the communication network and be further analyzed by project managers to better allocate tasks and staff.

The results of the SNA discussed above are for the combination of formal (contractual) and informal communications. However, social network studies have another category for studying the relationship between the environment (e.g., contractual relationship and conditions) and network structure that is beyond the scope of this work, but could be studied in the future. Future work could also explore the relationships between change order characteristics and social network attributes. Further information about change orders (e.g., cost and time impact, or type of change) could be taken into account when analyzing relationships between communication network characteristics and change orders. The effective use (pros and cons) and implementation of the proposed framework could be expanded by interviewing project participants, and further information about project participants (e.g., gender, age, experience in the industry) could also be taken into consideration.

4.7 Conclusion

Change order data were analyzed to assess the communication performance and centrality of project participants in a change-management network. Measuring communication through time-consuming content analysis to reveal the communication performance is difficult. This study measured centrality and performance using social network analysis. Project management may use the proposed framework and subsequent results to improve individual and team performance in the project change order process. This research contributes to the state-of-the-art by proposing an innovative use for a trusted methodology to discover social networks. Practitioners may use this framework to provide insight into relationships between characteristics of a communication network, or to quantitatively measure the performance of individual project members. Project managers can refer to these centrality measures during project execution to troubleshoot communication problems. Centrality measures can be used to understand performance and avoid delays and cost overruns.

4.8 Acknowledgments

This research work is funded by a Collaborative Research and Development Grant (CRDPJ 492657) from the Natural Science and Engineering Research Council. The authors also would like to acknowledge Dr. Mickey Richards for her assistance with manuscript editing and composition.

Chapter 5

Conclusions, Limitations, and Future Directions

5.1 Research Conclusions

This research outlines the development of a framework to enhance contract-related reporting and communication in construction. The framework functions as a systematic guideline to enhance documentation, communication, and sharing practices in both the planning and the execution and control phases of construction projects to fill existing research gaps in literature.

Chapter 2 develops a framework capable of automating the identification and extraction of reporting requirements and predicating their associated time and cost. The framework employs Natural Language Processing (NLP), Machine Learning (ML), and stochastic simulations to rapidly and efficiently retrieve requirements and quantify the time and costs associated with reporting—in turn providing necessary insights to streamline reporting workflows. To automate reporting requirement extraction, rule-based and ML-based classification methods were developed. Functionality and validity of both models were demonstrated using real contractual documents, and an accuracy of over 95% was observed. Then, numerical data regarding report preparation times and associated resources were used to predict the time and cost required to complete the reporting requirements detailed in the contractual documents. Input of these data into the Monte Carlo simulation model resulted in probability distributions that were validated by subject-matter experts, which confirmed that the simulated results were acceptable and were consistent with the outcomes observed in practice (a mean cumulative reporting duration and cost of 5083 hours and 304,939 CAD were observed, respectively).

Chapter 3 proposes a novel, reproducible framework for identifying areas of improvement in construction administration and quantitatively predicting their impact on project performance for enhanced decision-making. The framework employs data collection, value-stream mapping, and discrete-event simulation to stochastically model administrative processes and proposes performance metrics to quantitatively assess their performance. A case study was employed to demonstrate the applicability of both the framework and the metrics proposed, where both were found capable of reliably quantifying current and predicting future administrative performance of a construction contractor. Several performance improvements were observed in terms of resource utilization, process time, non-value-added time, and number of process steps—ultimately, the number of working days required was reduced by 31% from 489 days to 335 days.

Chapter 4 proposes a social network analysis-based approach to quantify and analyze the communication of documentation processes. The approach is focused on the communication of the participants of the reporting process, focusing on change order processes as one type of construction report. Project members involved in the change order process and their interactions were mined to assess the structural characteristics of the communication network at project, inter-organizational, change-order, and time-dependent levels. Communication data were extracted from a real project and used as a case study to demonstrate the applicability of the proposed approach. The results of the study demonstrate how individuals were engaged in and impacted the communication network, emphasizing the significance of effective communication in process-efficiency enhancement. The results quantitatively illustrated the communication network and highlighted the roles that key participants played. Key participants who influenced the social network were revealed at different levels of the study. The findings in this research can assist construction project managers in the development of appropriate staffing and task allocation strategies.

5.2 Academic Contributions

These research outcomes have resulted in several academic contributions:

- Automating contract-related reporting requirement identification and extraction to advance the contract review process.
- Advancement of text classification approaches using NLP techniques to provide a domain-specific and application-oriented text classification process.
- Provision of valuable insights and understanding regarding the prediction and analysis of the overhead costs and durations associated with contract-related reporting documentation.
- Introducing novel approaches for improving and streamlining contract-related reporting documentation in construction.
- Establishing generic performance indicators in construction administration to assess and quantify current and predict the future performance of contract-related administrative processes in construction.
- Integrating Lean principles and simulation techniques for administrative processes, which have typically been studied separately.
- Providing a foundation for the development of future Lean office strategies and tools in the

construction industry.

- Applying the social network analysis to discover and visualize communication in contract-related administrative processes in construction.
- Defining various communication indicators to measure and analyze communication performance of project members and to identify communication bottlenecks in the process, thereby enhancing the documentation process.

5.3 Industrial Contributions

Industrial contributions that have arisen out of collaborative research efforts with partner organizations include:

- Development of an automated tool to more efficiently identify reporting requirements and quantify the time and costs associated with report preparation. The proposed tool reduces the effort, time, and overhead costs expended by the multiple personnel involved in documentation process. A thorough and realistic understanding of contract reporting requirements promotes project teams to focus on establishing the best means, methods, pricing, and schedules for completing the proposed project.
- Providing practitioners with a better understanding of time, cost, and resource requirements, enabling the enhancement of contract negotiations, reporting workflow processes, and submittal procedures between clients and contractors, in turn increasing value for all project stakeholders.
- The use of multiple indicators to quantify process performance during the execution and control phases to provide decision-makers with the information required to take necessary

actions to streamline and smooth workflow and process design.

- Simulation, experimentation, and evaluation of the future behaviour of documentation processes to support managerial decision-making, thereby enhancing resource planning and allocation, reducing inefficiency and redundancy, and improving overall project outcomes.
- Providing indicators to better understand performance and to identify bottlenecks and misalignments in the process that impede project effectiveness. Using this approach on an ongoing construction project assists project managers to detect potential phenomena leading to communication problems and difficulties, thereby avoiding delays and cost overruns. Communication networks can be redesigned to involve the same participants more effectively by capitalizing on their skills, leading to improved process performance.

5.4 Research Limitations

Although the research findings in above chapters support the developed approaches, certain limitations of this research should be noted and explored.

- While the research framework is expected to be applicable—in its current form—to all contract-related documentation in construction sharing similar characteristics (e.g., mandated by contract and project specifications, carry information, flow through various process steps, and shared among various project participants), applicability of the framework to other documents types has yet to be confirmed using actual project data. Here, the sample application of the proposed framework was limited to the report documents (and their processes) of complex industrial projects, as access to other types of contract-related documentation was not available. Similarly, the quantification and streamlining of documentation workflows by focusing on people and communication was limited to construction change order reports.

- The extraction model in chapter 2 is developed and validated using one set of contract and project specification documents obtained from an oil and gas project. While the extraction module is expected to be applicable—in its current form—to all construction contracts with similar characteristics (e.g., terminology, document structure, and/or report structure), the development methodology described may need to be reapplied and revalidated for other contract types.
- The probability distributions used as input to the simulation model to predict time and cost must be investigated and discussed. Collecting and acquiring pertinent and correct information that accurately reflects the uncertainties associated with each variable is time-consuming. Certain construction companies do not even track the resources and time required from construction sites and administration offices to complete reporting requirements.
- The performance metrics proposed in chapter 3 may need to be adjusted depending on the nature of the administrative process under study. In addition, while proposed improvements can be used as a guide, improvements may not be applicable to all projects or organizations.
- The accuracy and reliability of the developed simulation model in Chapter 3 largely depends on the quality and availability of the data provided by the company.

5.5 Future Directions

This section reveals possible future directions based on this doctoral research work, which include:

- The development of a more holistic solution for contract documentation problems capable of providing seamless integration between clients and contractors, such as creating contracts using a structured-database approach.
- The research and development of methods for dealing with modifications or alternate arrangements during contract documentation.
- The exploration and development of simple, quick methods for collecting performance data of administrative processes in practice to help manage administrative processes.
- The development of approaches capable of quantitatively assessing potential influence of production variables on the performance of other variables, such as cost and quality of the process.
- The examination of the relationship between the environment (e.g., contractual relationship and conditions) and network structure in contract-related documentation processes.
- The investigation of relationships between report characteristics (e.g., type of report), project participant characteristics (e.g., gender, age, experience in the industry), and social network attributes.

References

Abdulmalek, F.A. and Rajgopal, J. (2007), "Analyzing the benefits of Lean manufacturing and value stream mapping via simulation: a process sector case study", *International Journal of Production Economics*, Vol. 107 No. 1, pp. 223–236.

AbouRizk, S. (2010), "Role of simulation in construction engineering and management", *Journal of Construction Engineering and Management*, Vol. 136 No. 10, pp. 1140–1153.

AbouRizk, S., Hague, S., Ekyalimpa, R. and Newstead, S. (2016), "Simphony: a next generation simulation modelling environment for the construction domain", *Journal of Simulation*, Taylor & Francis, Vol. 10 No. 3, pp. 207–215.

AbouRizk, S.M. and Halpin, D.W. (1992), "Statistical properties of construction duration data", *Journal of Construction Engineering and Management*, Vol. 118 No. 3, pp. 525–544.

Al Hattab, M. and Hamzeh, F. (2018), "Simulating the dynamics of social agents and information flows in BIM-based design", *Automation in Construction*, Vol. 92, pp. 1–22.

Al Qady, M. and Kandil, A. (2010), "Concept relation extraction from construction documents using natural language processing", *Journal of Construction Engineering and Management*, American Society of Civil Engineers, Vol. 136 No. 3, pp. 294–302.

Al Qady, M. and Kandil, A. (2013), "Document management in construction: Practices and opinions", *Journal of Construction Engineering and Management*, American Society of Civil Engineers, Vol. 139 No. 10, p. 06013002.

Al Qady, M. and Kandil, A. (2015), "Automatic classification of project documents on the basis of text content", *Journal of Computing in Civil Engineering*, Vol. 29 No. 3, p. 04014043.

Alnuaimi, A.S., Taha, R.A., Al Mohsin, M. and Al-Harthi, A.S. (2009), "Causes, effects, benefits, and remedies of change orders on public construction projects in Oman", *Journal of Construction Engineering and Management*, Vol. 136 No. 5, pp. 615–622.

AlSehaimi, A. and Koskela, L. (2008), "What can be learned from studies on delay in construction?", *16th Annual Conference of the International Group for Lean Construction*, International Group for Lean Construction, pp. 95–106.

Alves, T. da C., Pestana, A.C.V., Gilbert, E. and Hamzeh, F. (2016), "Lean principles for the management of construction submittals and RFIs", *Journal of Professional Issues in Engineering Education and Practice*, Vol. 142 No. 4, p. 05016004.

Atieh, A.M., Kaylani, H., Almuhtady, A. and Al-Tamimi, O. (2016), "A value stream mapping and simulation hybrid approach: application to glass industry", *The International Journal of Advanced Manufacturing Technology*, Vol. 84 No. 5–8, pp. 1573–1586.

Babalola, O., Ibem, E.O. and Ezema, I.C. (2019), "Implementation of Lean practices in the construction industry: a systematic review", *Building and Environment*, Vol. 148, pp. 34–43.

Ballard, G. and Howell, G. (2003), "Lean project management", *Building Research & Information*, Vol. 31 No. 2, pp. 119–133.

Barlow, G., Dew, C. and Woolley, P.D., Harriet. (2014), "Effective reporting for construction projects: increasing the likelihood of project success", *KPMG*, available at: https://assets.kpmg/content/dam/kpmg/pdf/2014/07/KPMG-PALS-13-Effective-reporting-for-construction-projects.pdf (accessed 12 November 2020).

Belayutham, S., González, V.A. and Yiu, T.W. (2016), "Clean–Lean administrative processes: a case study on sediment pollution during construction", *Journal of Cleaner Production*, Vol. 126, pp. 134–147.

Bergstra, J. and Bengio, Y. (2012), "Random search for hyper-parameter optimization", *The Journal of Machine Learning Research*, Vol. 13 No. 1, pp. 281–305.

Breiman, L. (2001), "Random forests", *Machine Learning*, Vol. 45 No. 1, pp. 5–32.

Bröchner, J. and Badenfelt, U. (2011), "Changes and change management in construction and IT projects", *Automation in Construction*, Vol. 20 No. 7, pp. 767–775.

Buckland, M. and Gey, F. (1994), "The relationship between recall and precision", *Journal of the American Society for Information Science*, Vol. 45 No. 1, pp. 12–19.

Butt, A., Naaranoja, M. and Savolainen, J. (2016), "Project change stakeholder communication", *International Journal of Project Management*, Vol. 34 No. 8, pp. 1579–1595.

Caldas, C.H., Soibelman, L. and Han, J. (2002), "Automated classification of construction project documents", *Journal of Computing in Civil Engineering*, Vol. 16 No. 4, pp. 234–243.

Charoenngam, C., Coquinco, S. and Hadikusumo, B. (2003), "Web-based application for managing change orders in construction projects", *Construction Innovation*, Vol. 3 No. 4, pp. 197–215.

Chen, J.C. and Cox, R.A. (2012), "Value stream management for Lean office—a case study", *American Journal of Industrial and Business Management*, Vol. 2 No. 2, pp. 17–29.

Chen, T. and Guestrin, C. (2016), "Xgboost: a scalable tree boosting system", presented at the Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794.

Chin, C. (2009), "Work-in-Process and Construction Project Information Flows", Vol. 1, presented at the Proceedings for the 17th Annual Conference of the International Group for Lean Construction, Table, pp. 257–266.

Chinowsky, P. and Taylor, J.E. (2012), "Networks in engineering: an emerging approach to project organization studies", *Engineering Project Organization Journal*, Vol. 2 No. 1–2, pp. 15–26.

Cooper, K.G. and Reichelt, K.S. (2007), "Project changes: Sources, impacts, mitigation, pricing, litigation, and excellence", *The Wiley Guide to Project Technology, Supply Chain & Procurement Management*, p. 347.

Costa, H.M., Silva, M.V.V., Mourão, A., Valente, C.P. and Laércio, F. (2013), "Redesigning administrative procedures using value stream mapping: a case study", *21st Conference of* *the International Group for Lean Construction*, International Group for Lean Construction, pp. 1049–1056.

Cox, I., Morris, J., Rogerson, J. and Jared, G. (1999), "A quantitative study of post contract award design changes in construction", *Construction Management & Economics*, Vol. 17 No. 4, pp. 427–439.

Dainty, A., Moore, D. and Murray, M. (2007), *Communication in Construction: Theory and Practice*, Routledge.

Das, M., Tao, X. and Cheng, J.C. (2020), "A Secure and Distributed Construction Document Management System Using Blockchain", presented at the International Conference on Computing in Civil and Building Engineering, Springer, pp. 850–862.

Desai, N., Pitroda, A. and Bhavsar, J. (2015), "A review of change order and assessing causes affecting change order in construction", *International Academic Research for Multidisciplinary*, Vol. 2 No. 12, pp. 152–162.

Dietrich, P., Eskerod, P., Dalcher, D. and Sandhawalia, B. (2010), "The dynamics of collaboration in multipartner projects", *Project Management Journal*, Vol. 41 No. 4, pp. 59–78.

Dogan, S.Z., Arditi, D., Gunhan, S. and Erbasaranoglu, B. (2013), "Assessing coordination performance based on centrality in an e-mail communication network", *Journal of Management in Engineering*, Vol. 31 No. 3, p. 04014047.

Dossick, C.S., Anderson, A., Azari, R., Iorio, J., Neff, G. and Taylor, J.E. (2014), "Messy talk in virtual teams: Achieving knowledge synthesis through shared visualizations”, *Journal of Management in Engineering*, Vol. 31 No. 1, p. A4014003.

East, E.W. and Love, D.R. (2011), “Value-added analysis of the construction submittal process”, *Automation in Construction*, Vol. 20 No. 8, pp. 1070–1078.

ElGindi, M. (2017), *User Friendly Progress Reporting System for Construction Projects*, *Dept. of Construction Engineering*, The American University in Cairo, 14 May, available at: http://dar.aucegypt.edu/handle/10526/5072 (accessed 12 November 2020).

El-Omari, S. and Moselhi, O. (2011), “Integrating automated data acquisition technologies for progress reporting of construction projects”, *Automation in Construction*, Vol. 20 No. 6, pp. 699–705.

ElSawy, I., Hosny, H. and Razek, M.A. (2011), “A neural network model for construction projects site overhead cost estimating in Egypt”, *International Journal of Computer Science*, Vol. 8 No. 3, pp. 273–283.

Erikshammar, J., Lu, W., Stehn, L. and Olofsson, T. (2013), “Discrete event simulation enhanced value stream mapping: an industrialized construction case study”, *Lean Construction Journal*, Vol. 10, pp. 47–65.

Fan, H. and Li, H. (2013), “Retrieving similar cases for alternative dispute resolution in construction accidents using text mining techniques”, *Automation in Construction*, Vol. 34, pp. 85–91.

Fernando, H., Hewavitharana, T. and Perera, A. (2019), "Evaluation of Electronic Document Management (EDM) systems for construction organizations", presented at the 2019 moratuwa engineering research conference (mercon), IEEE, pp. 273–278.

Fontanini, P.S.P., Picchi, F.A., Loureiro, S. and Lima Junior, O. (2008), "Simulating a construction supply chain– preliminary case study of pre-cast elements", *16th Annual Conference of the International Group for Lean Construction*, International Group for Lean Construction, pp. 371–380.

Forman, G. (2003), "An extensive empirical study of feature selection metrics for text classification", *Journal of Machine Learning Research*, Vol. 3 No. Mar, pp. 1289–1305.

Garrett, D.F. and Lee, J. (2010), "Lean construction submittal process—a case study", *Quality Engineering*, Vol. 23 No. 1, pp. 84–93.

Ghosh, S. and Burghart, J. (2019), "Lean construction: experience of US contractors", *International Journal of Construction Education and Research*, Vol. 17 No. 2, pp. 1–21.

Goienetxea Uriarte, A., Ng, A.H. and Urenda Moris, M. (2020), "Bringing together Lean and simulation: a comprehensive review", *International Journal of Production Research*, Vol. 58 No. 1, pp. 87–117.

Grefenstette, G. and Tapanainen, P. (1994), "What is a word, what is a sentence?: problems of Tokenisation", presented at the Proceedings of 3rd International Conference on Computational Lexicography, Research Institute for Linguistics, Hungarian Academy of

Sciences, Budapest, Hungary, pp. 79–87.

Hamzeh, F.R. (2011), "The Lean journey: implementing the last planner system in construction", *19th Annual Conference of the International Group for Lean Construction*, International Group for Lean Construction, pp. 379–390.

Hao, Q., Neelamkavil, J. and Shen, W. (2008), *Managing Changes in Construction*, Institute for Research in Construction, National Research Council Canada.

Hassan, F. ul and Le, T. (2020), "Automated requirements identification from construction contract documents using natural language processing", *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, Vol. 12 No. 2, p. 04520009.

Hastings, W.K. (1970), "Monte Carlo sampling methods using Markov chains and their applications", *Biometrika*, Vol. 57 No. 1, pp. 97–109.

Henderson, L.S., Stackman, R.W. and Lindekilde, R. (2016), "The centrality of communication norm alignment, role clarity, and trust in global project teams", *International Journal of Project Management*, Vol. 34 No. 8, pp. 1717–1730.

Hicks, B.J. (2007), "Lean information management: understanding and eliminating waste", *International Journal of Information Management*, Vol. 27 No. 4, pp. 233–249.

Hines, P. and Taylor, D. (2000), *Going Lean*, Lean Enterprise Research Centre, Cardiff, U.K.

Howell, G.A. (1999), "What is Lean construction-1999", *7th Annual Conference of the*

International Group for Lean Construction, International Group for Lean Construction.

Ibbs, C.W., Wong, C.K. and Kwak, Y.H. (2001), "Project change management system", *Journal of Management in Engineering*, Vol. 17 No. 3, pp. 159–165.

Ibbs, W. (2013), "Thinking about delay, disruption, and the cumulative impact of multiple changes", American Society of Civil Engineers, available at:https://doi.org/10.1061/(ASCE)LA.1943-4170.0000126.

Jafari, P., Al Hattab, M., Mohamed, E. and AbouRizk, S. (2021), "Automated extraction and time-cost prediction of contractual reporting requirements in construction using natural language processing and simulation", *Applied Sciences*, Vol. 11 No. 13, p. 6188.

Jafari, P., Mohamed, E., Lee, S. and Abourizk, S. (2020), "Social network analysis of change management processes for communication assessment", *Automation in Construction*, Vol. 118, p. 103292.

Jallow, A.K., Demian, P., Anumba, C.J. and Baldwin, A.N. (2017), "An enterprise architecture framework for electronic requirements information management", *International Journal of Information Management*, Vol. 37 No. 5, pp. 455–472.

Jeon, K., Lee, G. and Jeong, H.D. (2020), "Classification of the requirement sentences of the US DOT standard specification using deep learning algorithms", in Scheer, S. (Ed.), , presented at the Proceedings of the 18th International Conference on Computing in Civil and Building Engineering, Springer International Publishing, Cham, pp. 89–97.

Jeong, H.D., Gransberg, D. and Shrestha, K.J. (2015), "Framework for advanced daily work report system", *Mid-America Transportation Center*, available at: http://matc.unl.edu/assets/documents/matcfinal/Jeong_FrameworkforAdvancedDailyWorkReportSystem.pdf (accessed 12 November 2020).

Karim, A. and Adeli, H. (1999), "CONSCOM: An OO construction scheduling and change management system", *Journal of Construction Engineering and Management-Asce*, Vol. 125 No. 5, pp. 368–376.

Kemmer, S.L., Alves, T. da C.L., Macedo, M., Novaes, M. de V. and Barros Neto, J. de P. (2009), "Lean office at a construction company", *17th Annual Conference of the International Group for Lean Construction*, International Group for Lean Construction, pp. 43–52.

Kermanshachi, S., Thakur, R. and Govan, P. (2018), "Discovering the impact of late change orders and rework on labor productivity: a water treatment case study analysis using system dynamics modeling", presented at the Proceedings of the Construction Research Congress, New OrLeans, LA, USA, pp. 2–4.

Keyte, B. and Locher, D.A. (2004), *The Complete Lean Enterprise: Value Stream Mapping for Administrative and Office Processes*, CRC Press, Boca Raton, FL.

Ko, C.-H. and Li, S.-C. (2015), "Lean concurrent submittal review systems", *KSCE Journal of Civil Engineering*, Vol. 19 No. 3, pp. 478–484.

Koskela, L. (1992), *Application of the New Production Philosophy to Construction*, No. 72, Stanford University, Stanford, CA.

Koskela, L., Howell, G., Ballard, G. and Tommelein, I. (2002), “The foundations of Lean construction”, *Design and Construction: Building in Value*, Vol. 291, pp. 211–226.

Lee, G., Cho, J., Song, T., Roh, H., Jung, J., Chung, J., Yong, G., *et al.* (2020), “Construction Field Management Using a Popular Text Messenger”, in Scheer, S. (Ed.), , presented at the Proceedings of the 18th International Conference on Computing in Civil and Building Engineering, Springer International Publishing, Cham, pp. 971–979.

Lee, J., Yi, J.-S. and Son, J. (2019), “Development of automatic-extraction model of poisonous clauses in international construction contracts using rule-based NLP”, *Journal of Computing in Civil Engineering*, Vol. 33 No. 3, p. 04019003.

Lee, S. and Pena-Mora, F. (2007), “Understanding and managing iterative error and change cycles in construction”, *System Dynamics Review*, Vol. 23 No. 1, pp. 35–60.

Lee, S., Peña-Mora, F. and Park, M. (2005), “Quality and change management model for large scale concurrent design and construction projects”, *Journal of Construction Engineering and Management*, Vol. 131 No. 8, pp. 890–902.

Lee, S., Pena-Mora, F. and Park, M. (2006), “Reliability and stability buffering approach: Focusing on the issues of errors and changes in concurrent design and construction projects”, *Journal of Construction Engineering and Management*, Vol. 132 No. 5, pp. 452–464.

Levin, P. (2016), “Construction contract claims, changes, and dispute resolution”, *American*

Society of Civil Engineers, available at:https://doi.org/10.1061/9780784414293.

Liker, J.K. (2004), *Toyota Way: 14 Management Principles from the World's Greatest Manufacturer*, McGraw-Hill Education, New York, NY.

Lima, M., Rolim, L. and Alves, T. (2010), "Value stream mapping of the architectural executive design in a governmental organization", *18th Annual Conference of the International Group for Lean Construction*, Vol. 32, International Group for Lean Construction, pp. 415–424.

Luz, G.P., Tortorella, G.L., Narayanamurthy, G., Gaiardelli, P. and Sawhney, R. (2020), "A systematic literature review on the stochastic analysis of value streams", *Production Planning & Control*, Vol. 32 No. 2, pp. 1–11.

Macías-Jiménez, M.A., Acosta-Fontalvo, L.C. and Jiménez-Barros, M.A. (2019), "Document management practices in SMEs: An information management capability-based approach", *Records Management Journal*, Emerald Publishing Limited.

Manata, B., Miller, V., Mollaoglu, S. and Garcia, A.J. (2018), "Measuring Key Communication Behaviors in Integrated Project Delivery Teams", *Journal of Management in Engineering*, Vol. 34 No. 4, p. 06018001.

Manning, C. and Schutze, H. (1999), *Foundations of Statistical Natural Language Processing*, MIT press.

Manning, C.D., Schütze, H. and Raghavan, P. (2008), *Introduction to Information Retrieval*, Cambridge university press.

Monteiro, J., Alves, A.C. and do Sameiro Carvalho, M. (2017), “Processes improvement applying Lean office tools in a logistic department of a car multimedia components company”, *Procedia Manufacturing*, Vol. 13, pp. 995–1002.

Monteiro, M.F., Pacheco, C.C., Dinis-Carvalho, J. and Paiva, F.C. (2015), “Implementing Lean office: a successful case in public sector”, *FME Transactions*, Vol. 43 No. 4, pp. 303–310.

Moreno, J.L. (1960), “The sociometry reader”.

Morgan, A. (2010), *Does Poor Project Governance Cause Delays?*, available at: https://www.pwc.co.uk/assets/pdf/society-of-construction-law.pdf (accessed 6 January 2021).

Moselhi, O., Assem, I. and El-Rayes, K. (2005), “Change orders impact on labor productivity”, *Journal of Construction Engineering and Management-Asce*, Vol. 131 No. 3, pp. 354–359.

Motawa, I.A., Anumba, C.J., Lee, S. and Pena-Mora, F. (2007), “An integrated system for change management in construction”, *Automation in Construction*, Vol. 16 No. 3, pp. 368–377.

Naftanaila, I. and Mocanu, M.D. (2014), “Lean Office (LinOff) 2. Value Stream Mapping”, *Calitatea*, Vol. 15 No. 141, p. 75.

Ohno, T. (1988), *Toyota Production System: Beyond Large-Scale Production*, CRC Press, Boca Raton, FL.

Omar, T. and Nehdi, M.L. (2016), “Data acquisition technologies for construction progress tracking”, *Automation in Construction*, Vol. 70, pp. 143–155.

Padalkar, M. and Gopinath, S. (2016), “Six decades of project management research: Thematic trends and future opportunities”, *International Journal of Project Management*, Vol. 34 No. 7, pp. 1305–1321.

Park, H., Han, S.H., Rojas, E.M., Son, J. and Jung, W. (2010), “Social network analysis of collaborative ventures for overseas construction projects”, *Journal of Construction Engineering and Management*, Vol. 137 No. 5, pp. 344–355.

Peffers, K., Tuunanen, T., Rothenberger, M.A. and Chatterjee, S. (2007), “A design science research methodology for information systems research”, *Journal of Management Information Systems*, Vol. 24 No. 3, pp. 45–77.

Pestana, A.C.V. and Alves, T. da C. (2012), “Mapping the submittal process in a design-bid-build project”, *Construction Research Congress 2012: Construction Challenges in a Flat World*, American Society of Civil Engineers, Reston, VA, pp. 80–89.

Pestana, A.C.V., Alves, T. da C. and Barbosa, A.R. (2014), “Application of Lean construction concepts to manage the submittal process in AEC projects”, *Journal of Management in Engineering*, Vol. 30 No. 4, p. 05014006.

Porter, M. (1980), “An algorithm for suffix stripping program”, *Program: Electronic Library and Information Systems*, Vol. 14 No. 3, pp. 130–137.

Priyanka, H., Ramya, B. and Ashok, K. (2019), "Classification model to determine the polarity of movie review using logistic regression", *International Research Journal of Computer Science*, Vol. 6 No. 6, pp. 87–91.

Python. (2018), "Python 3.7.0", *Python Software Foundation*, available at: https://www.python.org/downloads/release/python-370/ (accessed 12 November 2020).

Rossiti, I., Serra, S.M. and Lorenzon, I.A. (2016), "Impacts of Lean office applications in the supply sector of a construction company", *24th Annual Conference of the International Group for Lean Construction*, International Group for Lean Construction, pp. 63–72.

Rother, M. and Shook, J. (2003), *Learning to See: Value Stream Mapping to Add Value and Eliminate Muda*, Lean Enterprise Institute, Boston, MA.

Russell, A.D. (1993), "Computerized daily site reporting", *Journal of Construction Engineering and Management*, Vol. 119 No. 2, pp. 385–402.

Russell, S. and Norvig, P. (2002), *Artificial Intelligence: A Modern Approach*.

Safapour, E. and Kermanshachi, S. (2018), "Identifying early indicators of manageable rework causes and selecting mitigating best practices for construction", *Journal of Management in Engineering*, Vol. 35 No. 2, p. 04018060.

Salama, D.M. and El-Gohary, N.M. (2016), "Semantic text classification for supporting automated compliance checking in construction", *Journal of Computing in Civil Engineering*, Vol. 30 No. 1, p. 04014106.

Sargent, R.G. (2010), “Verification and validation of simulation models”, *Proceedings of the 2010 Winter Simulation Conference*, Institute of Electrical and Electronics Engineers, Piscataway, NJ, pp. 166–183.

Sastre, R.M., Saurin, T.A., Echeveste, M.E.S., de Paula, I.C. and Lucena, R. (2018), “Lean office: study on the applicability of the concept in a design company”, presented at the DS 92: Proceedings of the DESIGN 2018 15th International Design Conference, pp. 643–654.

Sebastiani, F. (2002), “Machine learning in automated text categorization”, *ACM Computing Surveys (CSUR)*, Vol. 34 No. 1, pp. 1–47.

Serag, E., Oloufa, A., Malone, L. and Radwan, E. (2010), “Model for quantifying the impact of change orders on project cost for US roadwork construction”, *Journal of Construction Engineering and Management*, Vol. 136 No. 9, pp. 1015–1027.

Shash, A.A. and Habash, S.I. (2020), “Construction contract conversion: an approach to resolve disputes”, *Journal of Engineering, Project, and Production Management*, Vol. 10 No. 3, pp. 162–169.

Shiau, Y. and Wang, W. (2003), “Daily report module for construction management information system”, presented at the Proc. 20th Int. Symp. on Automation and Robotics in Construction, available at: https://www.researchgate.net/publication/267716061 (accessed 12 November 2020).

Shou, W., Wang, J., Wu, P., Wang, X. and Chong, H.-Y. (2017), “A cross-sector review on the use of value stream mapping”, *International Journal of Production Research*, Vol. 55 No. 13, pp. 3906–3928.

Shrestha, K.J. and Jeong, H.D. (2017), “Computational algorithm to automate as-built schedule development using digital daily work reports”, *Automation in Construction*, Vol. 84, pp. 315–322.

Shrestha, P.P. and Zeleke, H. (2018), “Effect of Change Orders on Cost and Schedule Overruns of School Building Renovation Projects”, *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, Vol. 10 No. 4, p. 04518018.

da Silva, I.B., Seraphim, E.C., Agostinho, O.L., Junior, O.F.L. and Batalha, G.F. (2015), “Lean office in health organization in the Brazilian Army”, *International Journal of Lean Six Sigma*, Vol. 6 No. 1, pp. 2–16.

Smit, M.C., Bond-Barnard, T.J., Steyn, H. and Fabris-Rotelli, I. (2017), “Email communication in project management: A bane or a blessing?”, *South African Journal of Information Management*, Vol. 19 No. 1, pp. 1–10.

Sun, M. and Meng, X. (2009), “Taxonomy for change causes and effects in construction projects”, *International Journal of Project Management*, Vol. 27 No. 6, pp. 560–572.

Sun, M., Sexton, M., Aouad, G., Fleming, A., Senaratne, S., Anumba, C., Chung, P., *et al.* (2004), “Managing changes in construction projects”, *EPSRC Industrial Report*, pp. 185–190.

Tapping, D. and Shuker, T. (2003), *Value Stream Management for the Lean Office: Eight Steps to*

Planning, Mapping, & Sustaining Lean Improvements in Administrative Areas, CRC Press, Boca Raton, FL.

Team, R.C. (2013), “R: A language and environment for statistical computing”.

Tixier, A.J.-P., Hallowell, M.R., Rajagopalan, B. and Bowman, D. (2016), “Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports”, *Automation in Construction*, Vol. 62, pp. 45–56.

Tommelein, I.D. (2015), “Journey toward Lean construction: pursuing a paradigm shift in the AEC industry”, *Journal of Construction Engineering and Management*, Vol. 141 No. 6, p. 04015005.

Turk, Z., Björk, B.-C., Johansson, C. and Svensson, K. (1994), “Document management systems as an essential step towards CIC”, presented at the CIB W78 Workshop on Computer Integrated Construction.

Turner, J.R. and Müller, R. (2004), “Communication and co-operation on projects between the project owner as principal and the project manager as agent”, *European Management Journal*, Vol. 22 No. 3, pp. 327–336.

Tyagi, S., Choudhary, A., Cai, X. and Yang, K. (2015), “Value stream mapping to reduce the lead-time of a product development process”, *International Journal of Production Economics*, Vol. 160, pp. 202–212.

Valieva, I., Voitenko, I., Björkman, M., Åkerberg, J. and Ekström, M. (2021), "Multiple machine learning algorithms comparison for modulation type classification based on instantaneous values of the time domain signal and time series statistics derived from wavelet transform", *Advances in Science, Technology and Engineering Systems*, Vol. 6 No. 1, pp. 658–671.

Wang, P., Mohamed, Y., Abourizk, S.M. and Rawa, A.T. (2009), "Flow production of pipe spool fabrication: simulation to support implementation of Lean technique", *Journal of Construction Engineering and Management*, Vol. 135 No. 10, pp. 1027–1038.

Wasserman, S. and Faust, K. (1994), *Social Network Analysis: Methods and Applications*, Vol. 8, Cambridge university press.

Witten, I.H., Frank, E., Hall, M.A. and Pal, C.J. (2016), *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, 4th ed., Vol. 31, Morgan Kaufmann, Boston.

Womak, J., Jones, D.T. and Roos, D. (1990), *The Machine That Changed the World*, Free Press, New York, NY.

Yokoyama, T.T., de Oliveira, M.A. and Futami, A.H. (2019), "A systematic literature review on Lean office", *Industrial Engineering & Management Systems*, Vol. 18 No. 1, pp. 67–77.

Yu, H., Tweed, T., Al-Hussein, M. and Nasseri, R. (2009), "Development of Lean model for house construction using value stream mapping", *Journal of Construction Engineering and Management*, Vol. 135 No. 8, pp. 782–790.

Zahraee, S.M., Esrafilian, R., Kardan, R., Shiwakoti, N. and Stasinopoulos, P. (2021), "Lean construction analysis of concrete pouring process using value stream mapping and Arena based simulation model", *Materials Today: Proceedings*, Vol. 42, pp. 1279–1286.

Zhang, J., Zi, L., Hou, Y., Deng, D., Jiang, W. and Wang, M. (2020), "A C-BiLSTM Approach to Classify Construction Accident Reports", *Applied Sciences*, Vol. 10 No. 17.

Zhang, L. and Ashuri, B. (2018), "BIM log mining: discovering social networks", *Automation in Construction*, Vol. 91, pp. 31–43.

Zhao, Z.Y., Lv, Q.L., Zuo, J. and Zillante, G. (2010), "Prediction system for change management in construction project", *Journal of Construction Engineering and Management*, Vol. 136 No. 6, pp. 659–669.

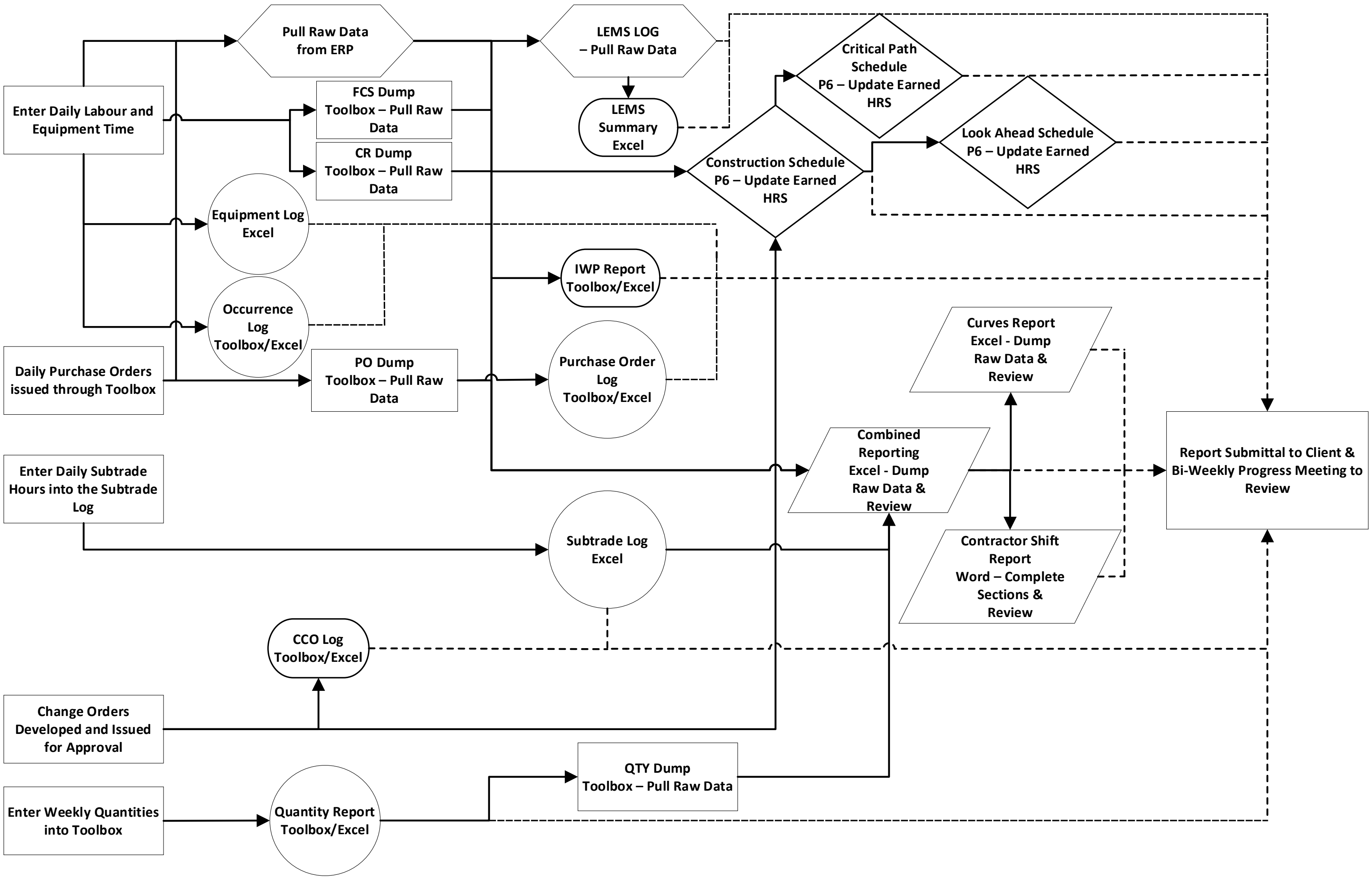
Zhong, B., Pan, X., Love, P.E., Ding, L. and Fang, W. (2020), "Deep learning and network analysis: Classifying and visualizing accident narratives in construction", *Automation in Construction*, Vol. 113, p. 103089.

Zhou, P. and El-Gohary, N. (2016), "Ontology-based multilabel text classification of construction regulatory documents", *Journal of Computing in Civil Engineering*, Vol. 30 No. 4, p. 04015058.

Appendix A. Project control reports flowchart

In the case study of Chapter 3, the administrative process associated with the preparation of project control reports for an oil and gas construction project led by a Canadian contractor in Alberta was considered to demonstrate the functionality of the proposed approach. To better understand and make a list of administrative tasks, the process flowchart is used to map the tasks and information flow. The process flowchart visualizes the flow of tasks and information that contractor has developed to collect required data, prepare, circulate, review, approve and submit multiple reports requested by the owner as per contractual agreement. As can be seen in the following figure, some reports are used as input to prepare other reports.

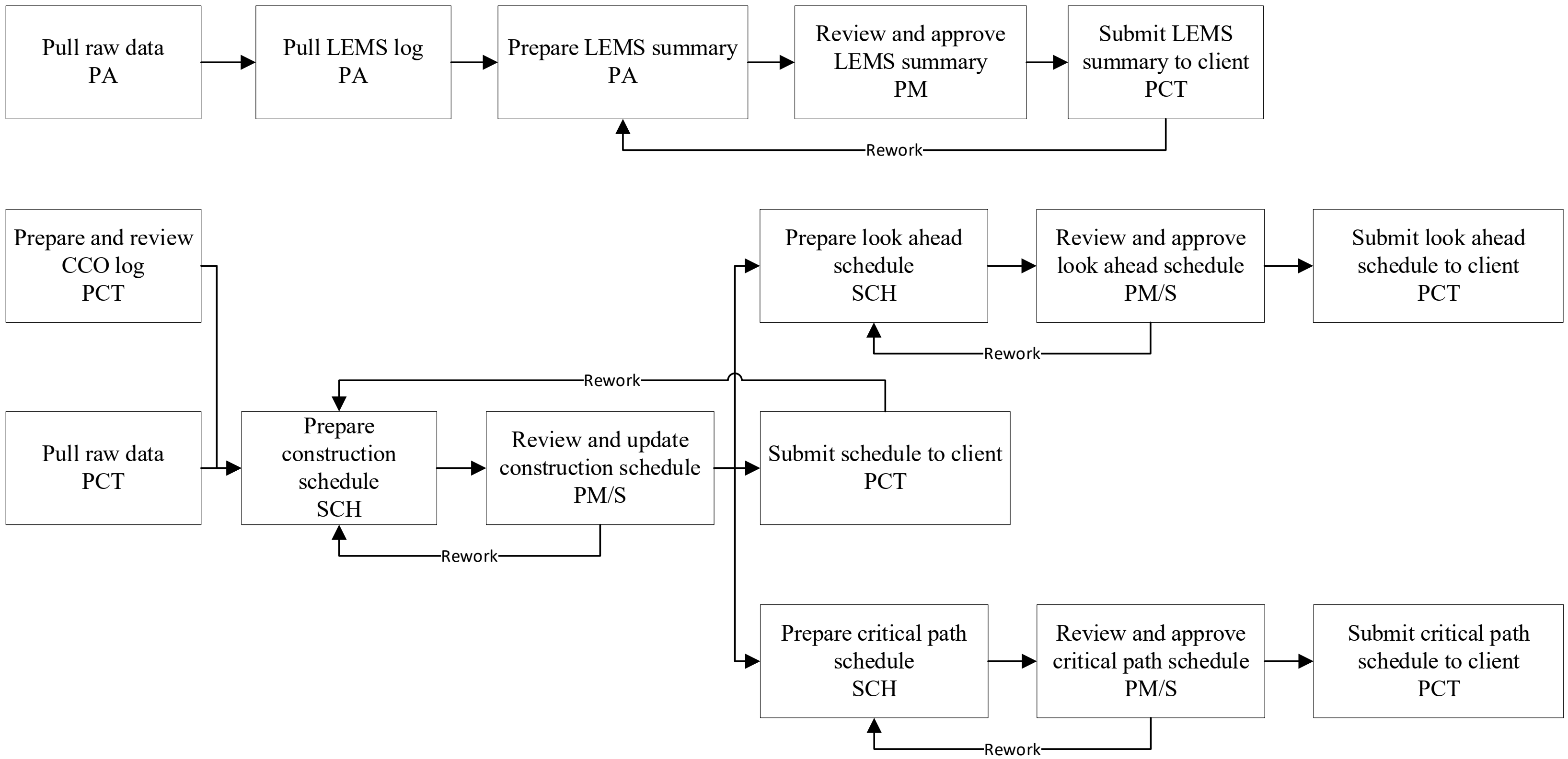
.

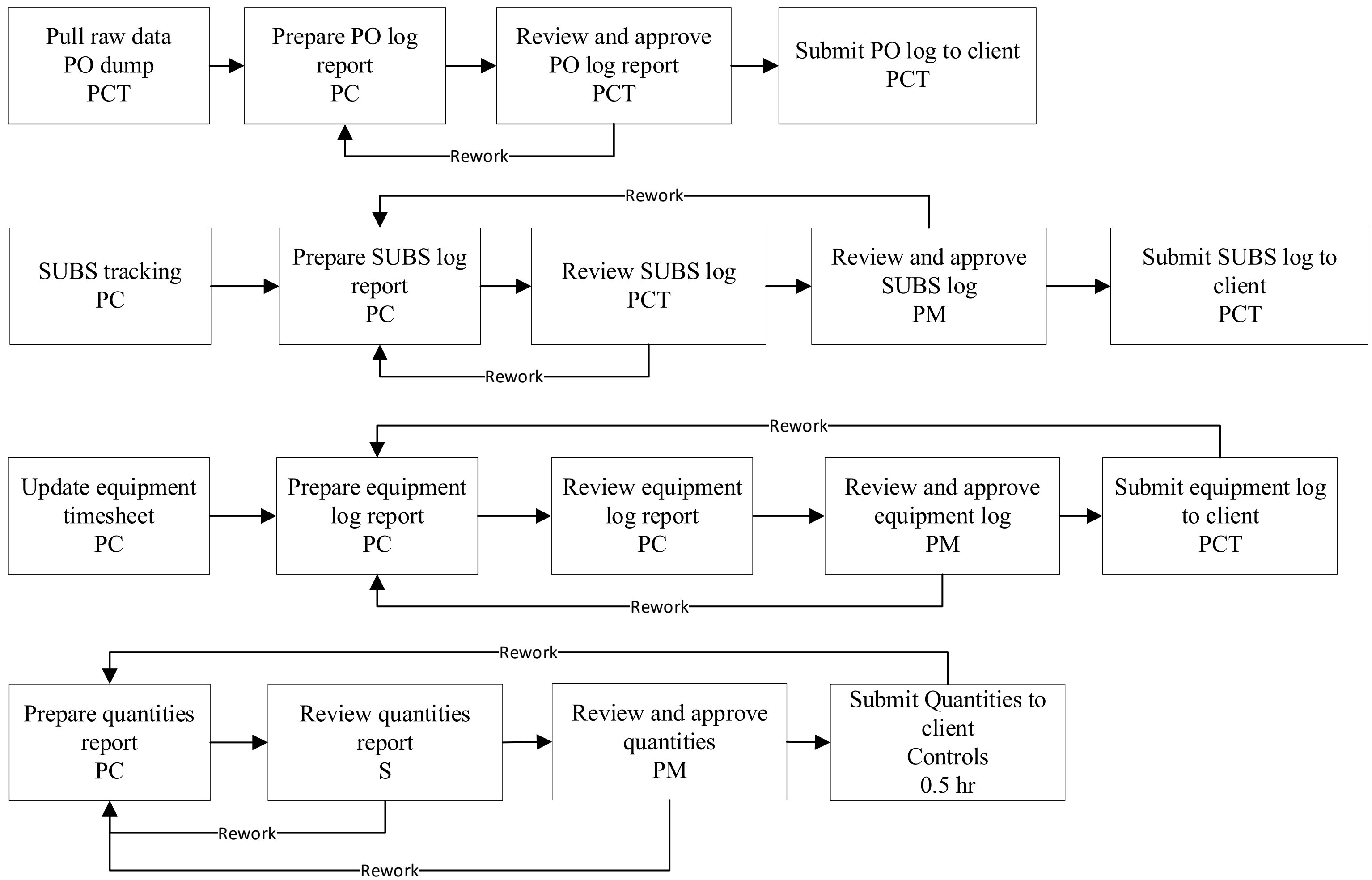

Pull Raw Data from ERP
LEMS LOG – Pull Raw Data
Enter Daily Labour and Equipment Time
FCS Dump Toolbox – Pull Raw Data
CR Dump Toolbox – Pull Raw Data
LEMS Summary Excel
Critical Path Schedule P6 – Update Earned HRS
Look Ahead Schedule P6 – Update Earned HRS
Construction Schedule P6 – Update Earned HRS
Equipment Log Excel
Occurrence Log Toolbox/Excel
IWP Report Toolbox/Excel
Daily Purchase Orders issued through Toolbox
PO Dump Toolbox – Pull Raw Data
Purchase Order Log Toolbox/Excel
Curves Report Excel - Dump Raw Data & Review
Combined Reporting Excel - Dump Raw Data & Review
Report Submittal to Client & Bi-Weekly Progress Meeting to Review
Enter Daily Subtrade Hours into the Subtrade Log
Subtrade Log Excel
Contractor Shift Report Word – Complete Sections & Review
CCO Log Toolbox/Excel
Change Orders Developed and Issued for Approval
QTY Dump Toolbox – Pull Raw Data
Enter Weekly Quantities into Toolbox
Quantity Report Toolbox/Excel

Appendix B. Value stream map of project control reports

In the case study of Chapter 3, after collecting the project information and required input data, current-state value stream maps were prepared and visualized using VSM. The value stream mapping is complicated consisting of 38 process steps with many links connecting the tasks and information flow. Here, value stream mapping of each report is separately visualized to simplify the complexity and to better understand the tasks and rework impact on each single report. The current-state maps are presented as follows:

.


Pull raw data
PA
Pull LEMS log
PA
Prepare LEMS summary
PA
Review and approve
LEMS summary
PM
Submit LEMS
summary to client
PCT
Rework
Prepare and review
CCO log
PCT
Prepare look ahead
schedule
SCH
Review and approve
look ahead schedule
PM/S
Submit look ahead
schedule to client
PCT
Rework
Rework
Pull raw data
PCT
Prepare
construction
schedule
SCH
Review and update
construction schedule
PM/S
Submit schedule to client
PCT
Rework
Prepare critical path
schedule
SCH
Review and approve
critical path schedule
PM/S
Submit critical path
schedule to client
PCT
Rework


Pull raw data
PO dump
PCT
Prepare PO log
report
PC
Review and approve
PO log report
PCT
Submit PO log to client
PCT
Rework
Rework
SUBS tracking
PC
Prepare SUBS log
report
PC
Review SUBS log
PCT
Review and approve
SUBS log
PM
Submit SUBS log to
client
PCT
Rework
Rework
Update equipment
timesheet
PC
Prepare equipment
log report
PC
Review equipment
log report
PC
Review and approve
equipment log
PM
Submit equipment log
to client
PCT
Rework
Rework
Prepare quantities
report
PC
Review quantities
report
S
Review and approve
quantities
PM
Submit Quantities to
client
Controls
0.5 hr
Rework
Rework

Appendix C. Scenario 4 and its impact on the process

In the case study of Chapter 3, after discussing the proposed process improvement solutions, experimental scenarios for the future-state map were developed. For each scenario, the impact of the proposed improvements on the design of the overall process, task durations, number of process steps, and the likelihood of rework was estimated. Scenario 4 had the lowest total duration and VA time of all experimented scenarios. The impact of the scenario 4 on the process tasks and associated durations is provided in the following table.

Impact of Scenario 4 on administrative process tasks and associated duration

Task	Name	Type[1]	Freq.	Duration (h)[2]	Resources[3]
1	Enter timesheets into SAP	NW	daily	U (2-3)	PC
2	Timecard keying	NW	daily	U (5-6)	PA
3	Subtrade tracking in the toolbox	VA	daily	U (1-2)	PC
4	Update equipment timesheet	NW	daily	Tri (0.75, 1, 1.25)	PC
5	Complete material receiving	NW	daily	Tri (1.5, 2, 2.5)	PA
6	Confirm completion of tasks	NW	daily	0.5	PC
7	Quantity entry	NW	daily	0.15-0.2	PC/PCT
8	Key subtrade invoices	NW	weekly	Tri (1-1.52)	PA
9	Post equipment timesheet	NW	weekly	Tri (0.5,0.75)	PC
10	Run LEMS	NW	shift	2	PA
11	Pull raw data from SAP and pull LEMS log	NW	shift	1	PA
12	Prepare LEMS summary	VA	shift	U (0.25,0.5)	PA
13	Review and approve LEMS summary	NW	shift	0.25	PM
14	Prepare and review CCO log	VA	shift	Tri (0.5, 1, 1.5)	PCT
15	Prepare construction schedule	VA	shift	Tri (2.5, 3, 4)	SCH
16	Review and update construction schedule	NW	Shift	U (1-2)	PM/S
17	Prepare look ahead schedule	VA	shift	Tri (0.5,1, 1.5)	SCH
18	Review and approve look ahead schedule	NW	shift	U (0.5-0.75)	PM/S
19	Prepare critical path schedule	VA	shift	Tri (0.5,1, 1.5)	SCH
20	Review and approve critical path schedule	NW	shift	U (0.5, 0.75)	PM/S

21	Prepare equipment log report AND review	VA	shift	Tri (1.5, 2, 2.5)	PC
22	Review and approve equipment log	NW	shift	U (0.5-1)	PM
23	Pull purchase order raw data and prepare	VA	shift	U (0.5-1)	PC
24	Review and approve purchase order log	NW	shift	1	PCT
25	Prepare subtrade log report	VA	shift	U (0.5-1)	PC
26	Review and approve subtrade log	NW	Shift	U (1-1.5)	PCT, PM
27	Prepare quantities report	VA	shift	U (0.5-1)	PC
28	Review and approve quantities report	NW	shift	U (1-1.5)	S, PM
29	Compile and prepare shift progress report	VA	shift	U (4-6)	PCT
30	Review and approve shift progress report	NW	shift	U (1-2)	PM
31	Submit bi-weekly reports	VA	shift	2	PCT

[1]VA: value-adding, NW: necessary waste, PW: pure waste
[2]Tri: triangular distribution, U: uniform distribution
[3]PC: project coordinator, PM: project manager, PA: project accountant, PCT: project control, S: superintendent, SCH: scheduler

The impact of the scenario 4 on the likelihood of revision occurrence, and the tasks required to be repeated to address the requested revision is presented in the following table.

Impact of Scenario 4 on rework

Rework (Revision)	Likelihood (%)	Requestor[1]	Affected Tasks
LEMS summary report	10	Client	12, 13
Construction schedule report	10	Client	15, 16
Construction schedule report 1st revision	10	PM	15, 16
Equipment log report	10	Client	21, 22
Equipment log report	5	PM	21, 22
Purchase order log report	5	PCT	23, 24
Subtrade log report	30	PCT, PM	23, 24
Quantities report	40	Client	27, 28
Quantities report 1st revision	40	S, PM	27, 28
Shift progress report	10	Client	29, 30

[1]PM: project manager, PA: project accountant, PCT: project control, S: superintendent