Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Engineering
Department of Mechanical and Industrial Engineering

Ingrid Renolen Borkenhagen
Jenni Sveen Olsen

Combining System Dynamics and Machine Learning for Predicting Safety Performance in Construction Projects

Master's thesis in Engineering and ICT
Supervisor: Nils Olsson
Co-supervisor: Antoine Rauzy
June 2023

NTNU
Norwegian University of
Science and Technology

Ingrid Renolen Borkenhagen
Jenni Sveen Olsen

Combining System Dynamics and Machine Learning for Predicting Safety Performance in Construction Projects

Master's thesis in Engineering and ICT
Supervisor: Nils Olsson
Co-supervisor: Antoine Rauzy
June 2023

Norwegian University of Science and Technology
Faculty of Engineering
Department of Mechanical and Industrial Engineering

NTNU
Norwegian University of
Science and Technology

Preface

This research is written as the master's thesis of our *Engineering and ICT* degree at the Norwegian University of Science and Technology (NTNU). The thesis is written in the spring of 2023 within the subject code *TPK4920 Project and Quality Management, Master's Thesis.* Further it counts for 30 credits.

Over the last years, we have both taken a selection of project management and machine learning courses. They have provided interesting and valuable knowledge that we wanted to utilize for our master's thesis. The project *Artificial Intelligence in Projects* was therefore a fitting choice as we could combine these two knowledge areas. Additionally, there is a rapid development within the machine learning field which we find very engaging as developers.

The selection of the project resulted in being a part of the sustainable value creation by digital predictions of safety performance in the construction industry (DiSCo) research project. Previously we have had limited knowledge on safety management in construction projects. This research has equipped us with new insights within this field, which we see great importance in. We have also identified a great potential for utilization of new, and better technology to further enhance safety performance in the construction industry. We hope that this thesis contributes with new knowledge and inspiration for further investigation of this domain.

Acknowledgement

We would like to thank all contributing parties to this thesis. Their invaluable support and guidance have been significant in our journey.

First and foremost, we would like to thank our supervisor Nils Olsson for great guidance during this semester. He has contributed with valuable insights and feedback, and we are truly greatful for his mentorship.

Further, we would like to thank our co-supervisor Antoine Rauzy and collaborating student Josefine Stiff Aamlid. Antoine's technical expertise and enthusiasm towards this research have greatly enriched our work. We would also like to thank Josefine for solid collaboration in the development of the system dynamics model.

Additionally, we would like to thank the rest of the DiSCo team for new perspectives and valuable discussions through our regular meetings. Their expertise has broadened our understanding and enriched the quality of our research.

At last, we would like to thank our friends, boyfriends and family for all the support. Our friends have contributed to some fantastic years in Trondheim, which also gives greater motivation for our studies. Thank you to our boyfriends and families for great support, encouragement and belief in our abilities.

Ingrid Borkenhagen

Ingrid Renolen Borkenhagen, June 2023

Jenni Sveen Olsen

Jenni Sveen Olsen, June 2023

Abstract

This thesis aimed to combine system dynamics and machine learning to give an early warning of construction projects with a high accident risk. The construction industry is highly accident prone and there is ongoing research on it's safety performance. Previous studies have focused on safety factors, system dynamics models as well as various machine learning predictions. For this thesis, a system dynamics model was developed in order to simulate construction projects. The simulation generated datasets which were utilized by machine learning models in order to predict safety performance.

The system dynamics model was developed based on theoretical findings. This thesis has an emphasis on the planning phase, as the model was developed in contribution with another master thesis *Integrating System Dynamics Modelling and Machine Learning to Improve Safety in Construction Projects* (Aamlid, 2023). The planning phase part of the model incorporated 53 indicators which influence each other and the construction phase. The system dynamics model was validated using extreme condition- and sensitivity tests, which showed reasonable trends towards the number of accidents. For each simulated project, all indicator data was saved into a dataset. This was done using two different accident rates; one for serious and one for fatal accidents. Consequently, two separate datasets were generated — one for serious accidents, which was balanced, and another for fatal accidents, which was highly imbalanced. Machine learning was applied to both datasets in order to predict safety performance. The datasets were preprocessed such that the features would only consist of planning phase data, with the target feature being if there had happened an accident or not. Five different machine learning models were utilized for this research.

The thesis revealed two primary findings. Firstly, the thesis successfully demonstrated the possibility of combining system dynamics and machine learning for safety predictions in cases where real project data is unavailable. Secondly, the results showed potential for separating the projects with higher risk of serious accidents and therefore give an early warning of poor safety performance. The serious accidents dataset yielded lower accuracies, yet higher recall values. However, the models struggled to identify fatal accidents. Due to the low values for the fatal accidents dataset, it was discussed how other safety measurements could be more applicable. Ultimately, the combination of system dynamics and machine learning has the potential to aid as decision support throughout construction projects and spread knowledge regarding safety performance.

Sammendrag

Denne oppgaven utforsket kombinasjonen av systemdynamikk og maskinlæring for å kunne gi en tidlig advarsel for byggeprosjekter som er spesielt utsatt for ulykker. Byggebransjen er preget av mye ulykker og det forskes på hvordan man kan forbedre sikkerhetsprestasjonen innen bransjen. Tidligere forskning har tatt for seg temaer som sikkerhetsfaktorer, systemdynamikk og prediksjon av sikkerhetsnivå. I denne oppgaven er det utviklet en systemdynamikk-modell for å simulere byggeprosjekter. Simulasjonen genererer datasett som benyttes for maskinlæring til å predikere sikkerhetsnivå.

Systemdynamikk-modellen ble utviklet basert på faglitteratur. Den ble utviklet i to deler; én for planleggingsfasen og én for byggefasen. Denne masteroppgaven tar for seg planleggingsfasen, mens byggefasen er implementert og beskrevet i *System Dynamics Modelling and Machine Learning to Improve Safety in Construction Projects* (Aamlid, 2023). Planleggingsmodellen inkluderte 53 indikatorer som påvirker hverandre og videre utvikling i byggefasen. Modellen ble validert gjennom ekstreme forhold- og sensitivitetstesting, og resultatene viste rimelige trender i forhold til antall ulykker. For hvert simulerte prosjekt ble indikatordataen lagret til et datasett. To forskjellige ulykkesrater ble brukt, en for alvorlige ulykker og en for fatale ulykker. Dette resulterte i to separate datasett; et for alvorlige ulykker og et for fatale ulykker. Datasettet for alvorlige ulykker var nokså balansert, mens datasettet for fatale ulykker var svært ubalansert. Maskinlæring ble anvendt på begge datasettene for å predikere sikkerhetsnivå. Datasettene ble forbehandlet slik at attributtene kun inneholdt data fra planleggingsfasen. Attributten som ble predikert var binær og beskrev om det hadde skjedd en ulykke eller ikke. Fem ulike maskinlæringsmodeller ble benyttet i studien.

Oppgaven avdekket to hovedfunn. For det første oppnådde oppgaven en vellykket integrering av systemdynamikk og maskinlæring for å predikere sikkerhetsnivå. For det andre viser resultatene potensial for å identifisere prosjekter med høy risiko for ulykker og dermed gi tidlig advarsel om lavt sikkerhetsnivå. Resultatene viste lav nøyaktighet, men høyere evne til å predikere de positive instansene korrekt for alvorlige ulykker. Maskinlæringsmodellene hadde imidlertid dårlig evne til å identifisere fatale ulykker. På grunn av de lave verdiene for å forutsi fatale ulykker ble det også diskutert hvordan alternative sikkerhetsmålinger kan være mer passende. Kombinasjonen av systemdynamikk og maskinlæring har potensial til å bidra som beslutningsstøtte og spre viktig kunnskap om sikkerhet i byggeprosjekter.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AI Artificial Intelligence.

ANN Artificial Neural Network.

BAR Blockchain Acceptance Rate.

BIM Building Information Modelling.

CPM Critical Path Method.

CSS Construction Safety Standards.

DiSCo Sustainable Value Creation by Digital Predictions of Safety Performance in the Construction Industry.

DT Decision Tree.

EPC Engineering, Procurement and Construction.

ESN Echo State Network.

FAR Fatal Accident Rate.

FN False Negative.

FP False Positive.

GA Genetic Algorithm.

GDPR General Data Protection Regulation.

HSE Health, Safety and Environment.

IoT Internet of Things.

LTI Lost-Time Injury Frequency Rate.

MLP Multi-Layer Perceptron.

MNOK Millions of Norwegian Kroners.

NTNU Norwegian University of Science and Technology.

PPV Positive Predictive Value.

RF Random Forest.

RNN Recurrent Neural Network.

RUO Reported Unwanted Occurrences.

SHA Safety, Health and Work Environment.

SMCS Smart Microclimate-Control System.

SVM Support Vector Machine.

SVR Support Vector Regression.

TN True Negative.

TP True Positive.

TPR True Positive Rate.

TRI Total Recordable Injury Frequency Rate.

VDT Virtual Design Team.

XGB Extreme Gradient Boosting.

1 Introduction

This thesis explores the utilization of system dynamics and machine learning to simulate construction projects and predict their safety performance. More specifically, the research will investigate whether machine learning predictions from simulated data can be used as an early warning for construction projects. The thesis also includes a theoretical background and multiple literature reviews to provide comprehensive context and background.

1.1 Background and Motivation

The construction industry is widely recognized as one of the most dangerous industries globally (Mohammadi, Tavakolan and Khosravi, 2018; Feng and Trinh, 2019; Li, Alburaikan and Fátima Muniz, 2023). This claim is supported by a number of sources, including a study done in 2022 by Mostue et al. that focuses on the high rate of work-related fatalities and accidents in Norway's construction industry. The construction industry is one of the most accident prone industries in Norway when looking at work related deaths and work related accidents (Mostue *et al.*, 2022). Nine fatalities were reported in Norway in 2021 (Mostue *et al.*, 2022). The average number of fatalities since 2012 is eight fatalities per year (Statistisk Sentralbyrå, 2022). Falls from heights and being struck by objects are among the most frequent causes of accidents in the construction industry (Sadeghi *et al.*, 2020; Aghaei, Asadollahfardi and Katabi, 2021). Additionally, incidents involving workers being squeezed/trapped are also common (Albrechtsen *et al.*, 2018). Out of the nine fatalities from 2021 three were from falling and four were from being squeezed/trapped (Mostue *et al.*, 2022).

Artificial intelligence (AI) is a rapidly developing technology with a wide range of potential applications that are currently starting to emerge across many industries (Regona *et al.*, 2022). Its potential applications are vast and continue to emerge as organizations recognize the transformative impact it can have. Machine learning, knowledge-based systems, computer vision, robotics, and optimization are examples of AI subfields. All of which have demonstrated their effectiveness in improving multiple aspects of industries. AI is currently transforming sectors like telecommunications, retail, and manufacturing (Abioye *et al.*, 2021). Machine learning, a key component of AI, enables systems to learn and make predictions or decisions without explicit programming. The different subfields of AI have been effectively implemented in different industries to boost profitability, efficiency, security, and safety (Abioye *et al.*, 2021). For organizations looking to improve their decision-making processes with quantitative support, using artificial intelligence and machine learning can be valuable. These advanced technologies have the capacity to evaluate enormous volumes of data and produce actionable recommendations, providing valuable insights for informed decision-making (Ramachandran *et al.*, 2022).

System dynamics leverages computer-based modeling and simulation methodologies to enable the study, analysis, and enhance decision-making in complex systems, utilizing mathematical modeling approaches (University of Bergen, 2022). In system dynamics projects, the objectives can vary. They may aim to develop theoretical understanding, put improvement measures in place, or often, pursue both goals simultaneously (Dangerfield, 2020). The origins of system dynamics are in business and organizational issues, but over time the application has spread to many fields such as healthcare, project management and economics (Dangerfield, 2020). These models have the capability to incorporate factors like stocks, flows, variables, and feedback loops. This allows for an examination of how changes in one part of the system can impact other components (Sweetser, 1999).

The construction industry is one the least digitized industries in the world (Abioye *et al.*, 2021). It's lack of digitization and highly manual nature makes project management more complicated and needlessly time-consuming. The insufficient degree of digitization in the construction industry has been associated with delays, inadequate performance in terms of quality and productivity, cost inefficiency, poor decision-making, and low performance in terms of health and safety (Abioye *et al.*, 2021). To address these challenges, the Sustainable value creation by digital predictions of safety performance in the construction industry (DiSCo) is a project at the Norwegian University of Science and Technology (NTNU) with the purpose of developing knowledge and methods for using AI in the early phases of construction projects to predict future safety performance in the production phase. It can therefore work as decision support for reducing the number of accidents (NTNU, n.d.). By leveraging technologies such as AI, the DiSCo project seeks to enhance safety outcomes and contribute to sustainable value creation in the construction industry.

1.2 Problem Description

The ultimate purpose behind this research is to improve the safety performance in the construction industry. For this purpose, the objective of the thesis is to utilize system dynamics and machine learning to give an early warning of projects with a high accident risk. The research will include developing a simulation model for the planning phase of construction projects. Data will be generated through this simulation and later combined with a range of machine learning models to predict accidents. The predictions will serve as a leading indicator, enabling warnings for accident-prone projects. Based on this objective, the following research questions have been formulated;

RQ1: How can system dynamics and machine learning be combined to predict safety performance in construction projects?

RQ2: What are the safety management implications of utilizing systems dynamics and machine learning in construction projects?

The first research question concerns the methodology and relation between the system dynamics model and machine learning predictions. It aims to explore the feasibility of combining these technologies to create a predictive model that can accurately forecast safety performance. The second research question concerns the results of combining system dynamics and machine learning, and what implications these give towards safety management in construction projects. This includes evaluating the extent to which safety performance can be predicted with these technologies and examining other implications resulting from this research.

1.3 Project Scope

The scope of this thesis is limited to the construction industry. However the methodology could be applied to other industries as well, such as the oil and gas or the manufacturing industry. Similarities can be drawn between these industries in terms of project data and health, safety, and environment (HSE) management practices. The scope of the thesis is additionally limited to the planning phase of construction projects. In the event that the results attains a high level of prediction accuracy, there is potential for its expansion into later project stages, including real-time predictions throughout the construction phase. This could allow for continuous monitoring and assessment of safety performance, enabling timely interventions and adjustments to minimize risks and ensure the welfare of workers. The system dynamics model is limited to construction projects of buildings. As the chosen contractual arrangement, the model is limited to projects with design-build contracts. Furthermore, the contract type was established as fixed price contracts.

1.4 Thesis Structure

Section 1 of this thesis serves as the introduction, exploring the background and motivation behind the research. It presents the research questions and project scope. In section 2 , the theoretical background of the thesis is presented. The first two subsections delve into project management and safety management in the construction industry. The following subsection explains system dynamics. Sections 2.4 and 2.5 provide important theoretical background on AI and machine learning in general, as well as specific information about the machine learning algorithms chosen for this thesis. The last subsection discusses the combination of system dynamics and machine learning. Section 3 details the methodology, including information about the thesis context, literature review, implementation of both the system dynamics and the machine learning models, and the evaluation of the method. Section 4 presents the results, including the tests conducted to validate the system dynamics model and the machine learning results for both serious accidents and fatal accidents. Section 5 is the discussion, reflecting on the system dynamics model, the machine learning models, and their respective results. The applicability of combining machine learning and system dynamics is also discussed. Lastly, section 6 presents the conclusion, relating the findings to the research questions, highlights the contributions of the thesis and suggesting possible further work.

2 Theoretical background

This section presents theoretical background regarding project management, safety management, system dynamics, AI in projects, machine learning and the combination of system dynamics and machine learning. The theory serves as a basis for research and development within this thesis.

2.1 Project Management

This section provides theoretical background regarding project management with an emphasis on construction projects. The roots of project management and several of its techniques can be traced back to construction projects, which served as the basis for *The Project Management Body of Knowledge* in 1987 (Project Management Institute, Inc., 2016). Project management can be defined as "the application of knowledge, skills, tools, and techniques to project activities to meet the project requirements" (Project Management Institute, Inc., 2017).

2.1.1 Project Definition

There are different ways to define a project. Tiltnes defines a project as a measure that has the character of a one-time initiative with a specific goal and a limited scope, carried out by a temporary organization within a time and cost frame (Tiltnes, 2015). According to the Project Management Institute "a project is a temporary endeavor undertaken to create a unique product, service or result" (Project Management Institute, Inc., 2017). Further, Rolstadås et al. describe how projects are often split into several tasks that are dependent on each other (Rolstadås *et al.*, 2020). Lastly, Eikeland describes projects as dynamic systems, as the tasks and demands for competency and resources change with time (Eikeland, 2001).

2.1.2 Construction Projects

Construction plays a big part of any advanced economy. It is estimated that construction accounts for 10% of the annual wealth generated in advanced economies, while around 20% in more recently industrialized nations (Winch, 2010). In Norway, the construction industry is the second biggest industry measured by value creation (Hansen, 2019).

Construction projects often result in a one-of-a-kind product (Project Management Institute, Inc., 2016). Hence, a construction project can be classified as a delivery project (Rolstadås *et al.*, 2020). Further, construction projects can be classified as either "green field" or "brown field" projects. "Green-field" projects are when the construction takes place in a new facility, while "brown-field" projects describe when projects are carried out in previously developed or abandoned sites (Project Management Institute, Inc., 2016). Construction projects are often performed in phases to be able to review and refine the design and strategy. Construction management, including planning and scheduling, cost management, risk management, document controls and forensic analysis, has been proven to increase the success of a project (Project Management Institute, Inc., 2016). Project success is often measured in time, cost and quality (Winch, 2010).

As construction work is organized in projects, there are fewer long term relations (Kvålshaugen and Swärd, 2018). This often applies to public construction projects, where the actors who offer the lowest price often win the project and the work is split into sub-contracts (Kvålshaugen and Swärd, 2018). The *project organization* describes the project owner's employees and all other actors, firms and people who hold significant roles within the project (Eikeland, 2001).

Even though construction projects are different and cannot be completely standardized, there are several similarities (Kvålshaugen and Swärd, 2018). Quality and complexity within construction projects are described below, in addition to typical actors, contractual arrangements, contract types and phases.

Quality

Quality can be defined as "degree to which a set of inherent characteristics of an object fulfills requirements" (ISO9000, 2015). In other words, quality describes in what degree the customers' requirements, needs and expectations are met (Lereim, 2013). Winch presents several aspects of construction project quality; product integrity regarding concepts, realisation and specification and conformance to requirements. He also states that quality in construction projects can be described as not product related, and that high quality processes are those that minimize risk of accidents (Winch, 2010). There are however different views on the quality term, and generally "high" quality will depend on the expectations of the user (Hansen, 2019). Further, different actors define quality depending on their competency (Hansen, 2019).

Complexity

There is no agreed definition of project complexity, but there is a common understanding that it involves more than just the size of the project (Johansen *et al.*, 2019). A conceptualization of project complexity done by Baccarini distinguishes between organization and technological complexity (Baccarini, 1996). Organizational complexity is determined by the number of organizational units, their tasks and relations (Johansen *et al.*, 2019). Technological complexity is defined based on the product itself (Rolstadås *et al.*, 2020). The International Centre for Complex Project Management describe complex projects as "undertakings for which traditional methods, practices and processes are inadequate in terms of scale, rate of change, heterogeneity, multiple pathways and ambiguous objectives" (International Centre for Complex Project Management, 2012). In addition, the stakeholders in complex projects are often very diverse (Johansen *et al.*, 2019).

Construction projects are often quite complex, and require different disciplines, technology and equipment (Kvålshaugen and Swärd, 2018). They are therefore prone to significant cost increases.

They vary widely on types, goals and solutions, and to each goal there can many many solutions (Project Management Institute, Inc., 2016). In addition, the projects occur in changing and complex environments (Project Management Institute, Inc., 2016). The project complexity affects the probability of success (Rolstadås *et al.*, 2020). The complexity is not necessarily described by the project itself, but can be subjective to the project organization's experiences (Rolstadås *et al.*, 2020).

Stakeholders

Stakeholders are defined as "individuals and organizations who are actively involved in the project, or whose interests may be positively or negatively affected as a result of project execution or successful project completion" (PMI Standards Committee and Project Management Institute, 1996). Owners and contractors are direct stakeholders in a construction project, as they are directly involved. There can also be many other stakeholders in a construction project. Stakeholders usually enter a project due to either geographical location or an invitation by the owner (Project Management Institute, Inc., 2016). Other direct project stakeholders include sponsors and suppliers, while indirect stakeholders can include regulatory agencies or authorities, professional associations, land owners and other project affected people. With this categorization, the indirect stakeholders describe those who are not directly involved, but can influence the project's execution.

Stakeholders can also be categorized as internal or external (Winch, 2010). Internal stakeholders are those who are in legal contact with the owner. External stakeholders are other stakeholders who also holds an interest in the project. Figure 1 shows a project stakeholder categorization by Winch, where it is distinguished between internal/external stakeholders as well as demand/supply side and private/public.

Internal stakeholders		External stakeholders	
Demand side	*Supply side*	*Private*	*Public*
Client	Architects	Local residents	Regulatory agencies
Financiers	Engineers	Local landowners	Local government
Client's employees	Principal contractors	Environmentalists	National
Client's customers	Trade contractors	Conservationists	Government
Client's tenants	Materials suppliers	Archaeologists	
Client's suppliers		Non-governmental organisations (NGO)	

Figure 1: Example of stakeholder categorization (Winch, 2010)

The project owner is typically involved in identifying and defining the project (Johansen *et al.*, 2019). In the context of other actors in the project, the owner is called the *client*. When the construction project is finished, the result becomes the owners property (Tiltnes, 2015). Even though the client can delegate tasks, they are responsible of overall decisions, defining the project scope and changes, as well as picking a contractor and approving the result (Tiltnes, 2015). Examples of clients are landowners, a municipality or an organization.

The executing parties in a project are hired by the client to transform resources into a result (Tiltnes, 2015). In construction projects these are often called the *contractors*. Contractors are responsible for the production in the project (Hansen, 2019). Examples of contractors can be firms offering masonry, carpeting, electrics, plumbing and/or digging. Sub-contractors are contractors being hired to perform a part of a main contractor's tasks. In addition to the main, technical contractors there are executing parties including architects and designers being hired by the client (Tiltnes, 2015). These work with detailed engineering in the design phase. This includes planning, forming and describing the product through sketches, models and descriptions (Hansen, 2019).

The users are also an important group within construction projects, as they are the ones who will utilize the building (Hansen, 2019). The user role will differ from project to project, depending on the project owner's purpose of the project. Additionally, one can see the users as a more complex aspect as they are not always one easily identifiable group (Olsson, Hansen and Blakstad, 2022).

Figure 2 shows the feedback loop between the main actors in a construction project. To create new facilities, it is important to learn from previous projects. This way, the client can make sure to meet the users' needs in a better way. This can be done by investigating previous user feedback to the client from similar projects. Further, the client will take this into account when working with the rest of the project organization (Winch, 2010).

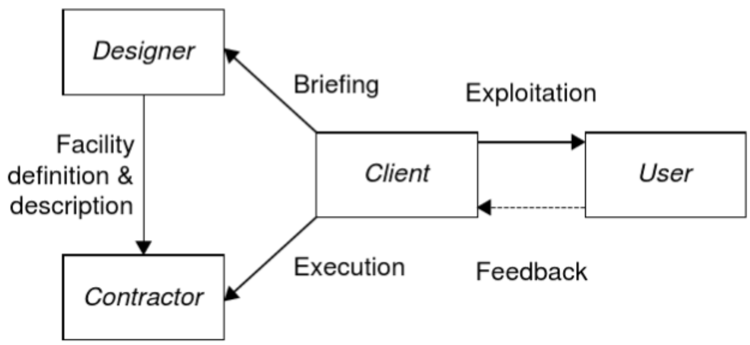


Figure 2: Feedback loop between actors (Winch, 2010)

Contractual Arrangements

Contractual arrangements describe the responsibilities of the client when hiring a contractor, especially regarding engineering and coordination. There are different types of arrangements, including design-bid-build contracts and design-build contracts. In design-bid-build contracts, the client is responsible for coordinating the engineering and implementation of the project (Direktoratet for forvaltning og økonomistyring, 2022b). As a result, the contractor is only responsible for the actual execution. Consultants and architects are often hired for the engineering. The second arrangement is design-build contract, also called engineering, procurement and construction (EPC) contract. In this model, the contractor is responsible for the engineering and implementation of the project (Hansen, 2019).

In Norway there has been established a set of standards regarding contractual arrangements. These are *NS8405* for design-bid-build contracts, *NS8407* for design-build contracts and *NS8401/02* for engineering consultants (Kvålshaugen and Swärd, 2018). These standards regulate circumstances such as financial conditions, solutions, responsibilities, safety and change management.

Contract Types

There are also different contract types regarding costs. Fee-based contracts are those who provide goods and services at a certain rate. These contracts are appropriate when the necessary resources may be easily identified but the amount is not yet known (Winch, 2010). Fixed-price contracts describe contracts where the sum is set for an agreed amount of work. These are either *lump-sum* contracts, where the price is fixed in the contract or *unit rate*, where the exact amount is not known until the work is done. Fixed-price contracts are appropriate when a big amount of information is known beforehand. Incentive contracts are a combination of fee-based and lump-sum. There are many variants of incentive contracts. However they have a common feature of trying to provide positive incentives through gainsharing.

Phases

The general phases in a construction project after initiating are planning, construction and delivery (Direktoratet for forvaltning og økonomistyring, 2022a). The planning consists of strategic definition, concept development, concept processing and detailed engineering (Tiltnes, 2015). These are described further in section 2.1.3.

The construction phase is the implementation of the construction work up until possibly the trial operation and the delivery (Direktoratet for forvaltning og økonomistyring, 2022a). One should follow up the contract to ensure that the building or facility conforms to the desired objectives. A kick-off meeting should be held in the beginning of the construction phase to provide a common understanding for all actors in the project (Direktoratet for forvaltning og økonomistyring, 2022a). There should also be developed a quality plan such that one can pay special attention to certain parts during the building process (Tiltnes, 2015).

After the building is completed it is time for a formal delivery. The delivery often consists of three phases; trial operation, takeover and potential complaints (Direktoratet for forvaltning og økonomistyring, 2022a). The latter is the follow-up of any deficiencies to make sure that rights are safeguarded.

After project delivery, Neste Steg additionally defines two more steps to the construction project life cycle; 1) use and administration and 2) liquidation (Tiltnes, 2015). The objective of the use and administration step is to secure a good technical and economic operation to satisfy the users' needs and intended effect (Tiltnes, 2015). Liquidation describes when the use of the building is being terminated. This occurs when the commercial activity ceases and the owner does not want to make a new investment. Then the building will typically be sold or demolished (Tiltnes, 2015).

2.1.3 Planning Construction Projects

Hansen states that one can consider a construction project as an interaction between several smaller processes of different characters. In this way, the project can be seen as the treatment of information which should take part in different activities which depend on and influence each other (Hansen, 2019). Generally, early phase planning consists of reciprocal and intensive dependencies, while the detailed engineering and building phase will be more sequential (Hansen, 2019).

There has recently been developed a Norwegian standard *NS3467* that describes all processes in construction projects. The standard is based on the framework *Neste Steg*. It aims to develop one common norm and industry language to better enhance cooperation between actors (Tiltnes, 2015). According to *Neste Steg* there are four stages before the construction begins; strategic definition, concept development, concept processing and detailed engineering (figure 3).

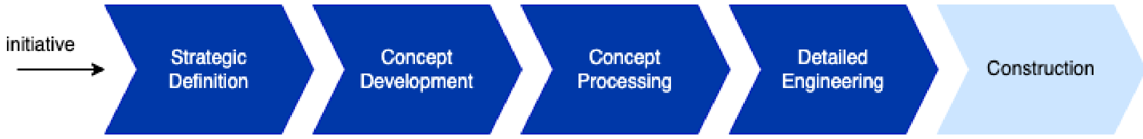


Figure 3: Planning stages in construction projects (inspired by Tiltnes, 2015)

Strategic Definition

The planning of a construction project typically begins with strategic planning. During this step, an idea is investigated and considered by the owner (Hansen, 2019). This step can either be initiated by the owner itself or by a user. The processes during this stage includes a gap analysis and the development of the business case (Tiltnes, 2015).

The gap analysis defines the gap between the current situation and the users' needs (Tiltnes, 2015). According to *Tidligfase i byggeprosjekter* needs will generally change over time (Iversen, Lilleland-Olsen and Woldseth, 2016). There are several types of needs; stakeholder needs, demand based needs and normative needs (Iversen, Lilleland-Olsen and Woldseth, 2016). Stakeholder needs can be when a building is not functionally fitted for the users. Demand based needs are often created when a business is growing. Normative needs are those that originate from laws and resolutions. The need that forms the idea in this stage is called the project triggering need (Iversen, Lilleland-Olsen and Woldseth, 2016). A stakeholder analysis should be conducted, as there are typically several needs (Iversen, Lilleland-Olsen and Woldseth, 2016). Stakeholder analyses uncover who gets influenced by a project and its results, who has a positive attitude towards the project and who is more skeptical (Lereim, 2013). This can display which stakeholders will have the most influence on the project. An associated stakeholder strategy should also be established (Tiltnes, 2015; Rolstadås *et al.*, 2020). This strategy serves as purpose to reduce the negative impact from the more skeptical stakeholders, while increasing the influence of the positive ones (Lereim, 2013).

The business case is a document that describes why the project should be realized (Lereim, 2013). It is developed by the owner and includes goals and ambitions, as well as a commercial framework (Tiltnes, 2015). Goals serve the purpose of specifying what one wishes to achieve and can be used afterwards to assess the degree of accomplishment (Hansen, 2019). The business case should list the objectives and reasons for project initiation (Project Management Institute, Inc., 2017). The management should examine viability as well as ethical concerns and consequences (Tiltnes, 2015). The goals that are defined in this stage should express the long-term business effects wanted to achieve after the project is realized (Iversen, Lilleland-Olsen and Woldseth, 2016). Further, they should match the strategic goals of the organization (Tiltnes, 2015; Iversen, Lilleland-Olsen and Woldseth, 2016; Hansen, 2019).

Rolstadås et al. presents the following model for business development (figure 4). First, business opportunities are identified based on a vision. Then the business case is developed based on opportunities and strategy. The business case should cover what is wanted to do, how it should be done and how much they expect to profit on the proposal. The business case makes it possible to estimate a business potential, which serves as the foundation for the establishment of a business plan (Rolstadås *et al.*, 2020). A project idea is developed based on the business case, which further will take part in the business plan.

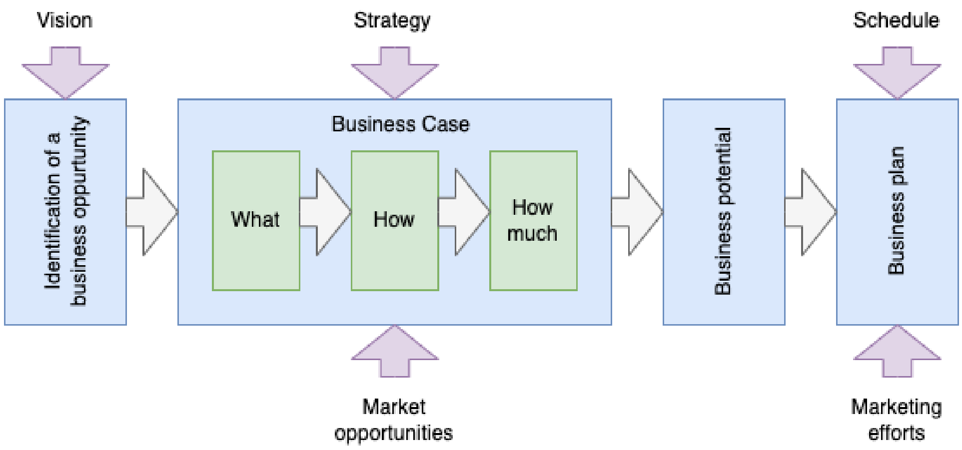


Figure 4: Business development (inspired by Rolstadås *et al.*, 2020)

During the strategic definition stage, an overall project plan should also be established (Direktoratet for forvaltning og økonomistyring, 2022a). The typical output of the strategic definition stage is an assessment of whether the project should be continued and a review of what possibilities there are for the project (Tiltnes, 2015).

Concept Development

The second stage of the planning phase is the concept development. The goal of this step is a determination of whether the project is feasible and to decide a concept for implementation (Hansen, 2019). Both the business case and the gap analysis from the previous stage is used as input for the activities in this stage (Tiltnes, 2015). There are several processes which should be executed during the concept development. These include developing a financial framework, alternative analyses and establishing an overall implementation strategy.

Prior to the alternative analysis there should be conducted a feasibility study (Iversen, Lilleland-Olsen and Woldseth, 2016). The first step should be to identify all solutions which meet the needs from the gap analysis. Then, each solution should be further analyzed on whether they are realistic and how well they meet the most important needs. The remaining solutions should be put in a prioritized list. This list is used as input for the alternative analysis.

An alternative analysis should assess various solutions including the 0-solution, being the current situation, against the project's objectives and requirements (Direktoratet for forvaltning og økonomistyring, 2022a). The analysis is both quantitative and qualitative (Iversen, Lilleland-Olsen and Woldseth, 2016). The purpose is choosing the best possible solution for the owner (Iversen, Lilleland-Olsen and Woldseth, 2016). The three steps of the alternative analysis are;

1. expanding the descriptions of the different alternatives
2. conducting a cost benefit analysis
3. comparing other effects such as safety, flexibility and environment

The analysis should conclude with a concept recommendation based on the measures above (Iversen, Lilleland-Olsen and Woldseth, 2016). The final output of this activity is the concept documents.

A financial framework should be established on a higher level. It should include construction costs, budget and risk assessments (Direktoratet for forvaltning og økonomistyring, 2022a). From the user perspective, a functional plan together with sketches and illustrations should be developed (Tiltnes, 2015). Once the concept and sketches are available, an overall implementation strategy should be established (Direktoratet for forvaltning og økonomistyring, 2022a). In addition to the deliveries above, the business case should be updated with the new information.

From a management perspective, the concept development step includes expanding the project plan, establishing a procurement strategy, preparing for contract- and organization structure and establishing a building information modelling (BIM) strategy (Tiltnes, 2015). A BIM strategy should contain information on how BIM should be used in the project, how it should be organized and what standards should be used (Hansen, 2019). A plan for communication and information handling should also be established during this stage. The concept development stage is also often where the management document is first established. It contains parts from the processes above, typically including plans, goals, general conditions, organizing, implementation strategy and project plan. The final step of this stage is deciding whether one should continue with the chosen concept (Tiltnes, 2015; Iversen, Lilleland-Olsen and Woldseth, 2016).

Concept Processing

Next, the chosen concept needs to be further developed. The goal of the concept processing stage is to analyze more in detail such that the final decisions regarding implementation and finances are made on the right basis (Tiltnes, 2015). Relevant inputs from previous stages include concept documents, overall implementation strategy, sketches and illustrations, functional capacities as well as the management documents.

During this stage the owner should make a financial plan, determine the contract- and organizational structure and update the business case for the project (Tiltnes, 2015; Iversen, Lilleland-Olsen and Woldseth, 2016). A sketch project takes place, which is the calculations and sketches that lay the foundation for architecture and design (Direktoratet for forvaltning og økonomistyring, 2022a). Further development of the sketch project is the preliminary project, which contains functional and physical structure (Direktoratet for forvaltning og økonomistyring, 2022a). An implementation strategy should be developed further, sketches of the building method created and technical constraints considered (Tiltnes, 2015). The implementation strategy contains contractual arrangements and contract type (Direktoratet for forvaltning og økonomistyring, 2022a). Towards the public a regulation plan should be made or updated, one should apply for a general permission and perform a risk assessment (Tiltnes, 2015). A regulation plan is a document to manage the development of a certain area (Hansen, 2019). It is formed by the owner and needs to be approved by the local authorities. A general permission covers the frame of the building, such as size, volume, environmental aspects, looks and neighbours (Hansen, 2019). The risk assessment focuses on identifying the potential opportunities and risks (Johansen *et al.*, 2019).

Other processes within this stage include updating the procurement strategy and the project plan, defining roles associated with communication and information handling, verifying BIM strategy and developing a safety, health and work environment (SHA) plan (Tiltnes, 2015). Before construction or installation work begins, a written SHA plan that outlines how the project's risk factors will be managed is required by the Construction Client Regulations § 7 (Arbeidstilsynet, 2023). When finishing the concept processing stage, the final decision on whether the project should be implemented is made.

Detailed Engineering

The final stage before the construction begins is the detailed engineering, also called the design phase. The purpose of this step is to develop a high quality foundation with enough details for a safe and correct construction phase (Tiltnes, 2015). The design phase is executed by specialists within their field, and relevant experience plays a big role (Lereim, 2013). Relevant inputs from previous stages include the decided concept (concept documents) and the contractual arrangement. Another input is information regarding suppliers' systems and products (Tiltnes, 2015). During this stage, a construction plan and updated business case should be made by the owner. The executive parties should establish a production base, with measures on both time, quality and implementation. Drawings and descriptions based on the preliminary project should be detailed and finalized (Direktoratet for forvaltning og økonomistyring, 2022a). Most documents are defined down to a three-digit level according to the building component table (Direktoratet for forvaltning og økonomistyring, 2022a). The BIM should be fully completed. A building application should be sent to the associated municipality (Tiltnes, 2015). The SHA plan and procurement strategy should also be updated.

To summarize, the four main stages during planning are strategic definition, concept development, concept processing and detailed engineering. Despite the fact that projects live in an uncertain world (Rolstadås *et al.*, 2020), a big amount of planning is usually done before the construction phase begins. Planning plays a big role for the execution, and is particularly important as projects are only given one chance (Lereim, 2013). According to several case studies, planning in the early phases is shown to be one of the most important success factors for construction projects (Hussein, 2016).

2.2 Safety Management in the Construction Industry

This section concerns safety management within the construction industry. As presented in the introduction (section 1.1), the construction industry suffers from low safety performance. This leads to an increasing amount of research on the area. Theory regarding different safety measurements as well as a literature review on safety factors are provided.

2.2.1 Measuring Safety Performance

Safety performance indicators measure an organizations ability to control risk of accidents (Kjellen and Albrechtsen, 2017). Indicators are commonly used when direct measurements are too intricate. Sultana, Andersen and Haugen describe an indicator as a "measurable representation of the aspect of reality" (Sultana, Andersen and Haugen, 2019). For safety performance, one can distinguish between *leading* and *lagging* indicators.

Leading indicators is a term well known within the economy field, utilized by business economists to forecast the course of the economy in the near future (Stock and Watson, 2008). For safety purposes Kjellen and Albrechtsen define leading indicators as those which "predict future developments in safety performance, that is, they change before the safety performance has changed" (Kjellen and Albrechtsen, 2017). Lagging indicators on the other hand describe safety performance based on incidents that have already happened. They change after an activity's safety performance has changed.

Loss-based safety performance indicators are categorized as lagging indicators. Two common loss-based indicators are the lost-time injury (LTI) frequency rate and total recordable injury (TRI) frequency rate. The LTI rate describes the number of lost-time injuries per one million hours of work (Rausand, 2013). A lost-time injury can be defined as an injury that happened at work which results in the worker not being able to meet on the next shift (Kjellen and Albrechtsen, 2017). It used to be the most common measurement of safety performance. The indicator has however gained some negative attention as it can be considered as easy to manipulate. An example would be letting the injured person do "lighter" work the next day instead of being absent. The TRI rate describes the total number of recordable incidents. This term contains a much larger spectrum of accidents, as recordable incidents include all injuries at work resulting in lost time, medical treatment (not first-aid), restricted work and fatalities. This makes the TRI rate harder to manipulate than the LTI rate. Days since last LTI or TRI are also known measurements for safety performance. They do however not take into account the size of the company.

Other loss-based safety performance indicators include accumulated number of recordable injuries, severity rate (S-rate), average number of days lost, fatal accident rate (FAR) and accumulated number of fatalities (Kjellen and Albrechtsen, 2017). Accumulated number of recordable injuries measure the injuries since the beginning of an activity or project. Similarly to the days since last LTI or TRI, this indicator does not take into account the size of the company, and is therefore not as suited for when companies increase or decrease in number of employees. The S-rate is the number of working days lost as a result of lost-time injuries per one million hours of work (Kjellen and Albrechtsen, 2017). Fatalities and permanent disabilities corresponds to 7500 workdays. The S-rate indicator has been criticized for being dominated by long sick leaves. However, it is less sensitive to inaccuracies than the LTI and TRI. Another measurement is the average number of lost days, which is the S-rate divided by the LTI rate. The FAR, which describes the fatal accident rate, measures the number of fatalities per 100 million working hours within a defined population (Rausand, 2013). It is commonly used to compare different industries. The FAR is however rarely used within companies due to the low number of fatalities, as it is easier for the company to count the number directly (Kjellen and Albrechtsen, 2017). At last, the accumulated number of fatalities measures the accumulated number since the beginning of an activity or a project.

There are also process-based safety indicators. These include measures of both deviations and incidents. Examples are percentage compliance with regulations, percentage correct behaviour, number of incidents per period and number of reported unwanted occurrences (RUO) (Kjellen and Albrechtsen, 2017). The latter can be measured per employee and year.

There are many different ways of measuring safety performance. As mentioned above, these indicators have their strengths and weaknesses. Therefore Kjellen and Albrechtsen recommend using a combination of several indicators for safety measurements.

2.2.2 Safety Factors

There has been a large amount of research on what causes bad safety performance and why accidents occur. In this thesis, these are called *safety factors*. This section will describe previous studies and provide information on the factors that has occurred the most within this research field.

Khalid, Sagoo and Benachi performed an empirical study in order to develop a framework for improved safety performance in construction projects (Khalid, Sagoo and Benachir, 2021). The literature findings revealed 63 factors, which were later clustered into six groups. Mohammadi, Tavakolan and Khosravi conducted a qualitative content analysis to retrieve safety factors (Mohammadi, Tavakolan and Khosravi, 2018). It resulted in 13 factors. In similarity to Khalid, Sagoo and Benachi they proposed a framework based on these. The framework contains a hierarchy of the safety factors on several levels. Usukhbayar did a literature review on safety factors which resulted in 58 factors (Usukhbayar and Choi, 2020). They were grouped into 13 groups using adopting factor analysis methodology. Muñoz-La Rivera, Mora-Serrano and Oñate performed a systematic literature review, which led to the identification and classification of 100 factors (Muñoz-La Rivera, Mora-Serrano and Oñate, 2021). In addition, Yap and Lee did a literature review and questionnaire (Yap and Lee, 2020). This resulted in eight main factors. Abas et al. also conducted a literature review (Abas *et al.*, 2020).

Further, Asilian-Mahabadi et al. investigated factors contributing to unsafe work behaviours using field observations, interviews and focus group discussions (Asilian-Mahabadi *et al.*, 2018). This resulted in fourteen themes within four categories. Sukamani and Wang developed a model based on a questionnaire to identify factors of critical accidents (Sukamani and Wang, 2020). At last, Man et al. investigated the effect of factors on risk taking behaviour in construction projects (Man *et al.*, 2021). They used a combination of a questionnaire and structural equation modeling.

The literature findings revealed the following safety factors;

Organizational factors

Safety culture is an organizational factor that holds significant importance within a company (Khalid, Sagoo and Benachir, 2021; Muñoz-La Rivera, Mora-Serrano and Oñate, 2021). Mohammadi, Tavalokan and Khosravi mention attitudes, feedback, safety involvement and effort as safety factors (Mohammadi, Tavakolan and Khosravi, 2018). Asilian-Mahabadi found that the safety climate on both the client and contractor side have a big impact (Asilian-Mahabadi *et al.*, 2018). Additionally, attitude and motivation is highlighted in their study. Attitude was also mentioned as a safety factor by Man et al. (Man *et al.*, 2021).

Safety performance is also significantly impacted by economics (Khalid, Sagoo and Benachir, 2021; Muñoz-La Rivera, Mora-Serrano and Oñate, 2021). Mohammadi, Tavalokan and Khosravi mention project cost, contract price and project size as influencing factors (Mohammadi, Tavakolan and Khosravi, 2018). They also state that the company's revenue and costs are important factors. According to Asilian-Mahabadi, the safety performance is influenced by economic conditions (Asilian-Mahabadi *et al.*, 2018).

Another aspect covered by previous studies is the structure within the project organization. The involvement and size of subcontractors are mentioned as factors by several studies (Mohammadi, Tavakolan and Khosravi, 2018; Muñoz-La Rivera, Mora-Serrano and Oñate, 2021). In addition, a factor could be whether safety is included when selecting contractors (Sukamani and Wang, 2020). Other organizational factors which are presented in the articles are stakeholder management (Khalid, Sagoo and Benachir, 2021) and quality (Khalid, Sagoo and Benachir, 2021; Muñoz-La Rivera, Mora-Serrano and Oñate, 2021).

Managerial Factors

Managerial factors such as safety management systems, safety investments and risk assessments are found to be important regarding safety performance. The effectiveness of safety management systems are considered a contributing factor by several studies (Mohammadi, Tavakolan and Khosravi, 2018; Khalid, Sagoo and Benachir, 2021; Muñoz-La Rivera, Mora-Serrano and Oñate, 2021). Safety investments have also been found to impact safety performance (Mohammadi, Tavakolan and Khosravi, 2018; Khalid, Sagoo and Benachir, 2021; Muñoz-La Rivera, Mora-Serrano and Oñate, 2021). Additionally, risk assessment is a factor that was uncovered by Khalid, Sagoo and Benachir (Khalid, Sagoo and Benachir, 2021).

Further, supervision was highlighted in several studies (Asilian-Mahabadi *et al.*, 2018; Khalid, Sagoo and Benachir, 2021). Usukhbayar and Choi point out how frequency of inspections can affect the safety performance (Usukhbayar and Choi, 2020). Abas et al. also highlight how the implementation of safety inspections are important towards safety management (Abas *et al.*, 2020).

Legislative Factors

Safety policies and rules are legislative factors that affect safety performance (Usukhbayar and Choi, 2020; Khalid, Sagoo and Benachir, 2021; Muñoz-La Rivera, Mora-Serrano and Oñate, 2021). The compliance of these rules is listed as a factor by several studies (Mohammadi, Tavakolan and Khosravi, 2018; Khalid, Sagoo and Benachir, 2021; Muñoz-La Rivera, Mora-Serrano and Oñate, 2021). In connection to this, Abas et al. proposes "safe systems of work", being safety policies, as a factor (Abas *et al.*, 2020). Sukamani and Wang proposes "poor enforcement of safety rules and regulations by the government agencies" as a factor (Sukamani and Wang, 2020).

Environmental Factors

The reviewed articles demonstrate that the project's environment significantly influences its safety performance (Abas *et al.*, 2020; Sukamani and Wang, 2020; Yap and Lee, 2020). Factors such as weather (Usukhbayar and Choi, 2020; Khalid, Sagoo and Benachir, 2021) and safety hazards (Khalid, Sagoo and Benachir, 2021) can contribute to poorer safety. Lack of or poorer equipment is also a factor that contributes to a worse working environment (Mohammadi, Tavakolan and Khosravi, 2018; Abas *et al.*, 2020; Usukhbayar and Choi, 2020). Further, Sukamani and Wang found incorrect and defective tools to influence the safety performance of a project (Sukamani and Wang, 2020).

Personnel Factors

There are several personnel factors that are discovered in the reviewed studies. One factor is the attitude and motivation of the worker (Mohammadi, Tavakolan and Khosravi, 2018; Khalid, Sagoo and Benachir, 2021). Further the workers education, experience and training are listed as factors that influence their safety performance (Usukhbayar and Choi, 2020; Sukamani and Wang, 2020; Usukhbayar and Choi, 2020; Yap and Lee, 2020; Khalid, Sagoo and Benachir, 2021). This was also covered in the factor "contractor competency" by Asilian-Mahabadi et al. (Asilian-Mahabadi *et al.*, 2018). Safety training was listed as one of the most influential factors by Man et al. (Man *et al.*, 2021). Moreover, age is mentioned as a factor in several studies (Mohammadi, Tavakolan and Khosravi, 2018; Usukhbayar and Choi, 2020; Yap and Lee, 2020).

Another factor that has been revealed through the studies is pressure. This includes production pressure and schedule delays (Mohammadi, Tavakolan and Khosravi, 2018). Rework is also mentioned as a productivity aspect (Mohammadi, Tavakolan and Khosravi, 2018, Muñoz-La Rivera, Mora-Serrano and Oñate, 2021). Furthermore, overtime is a factor discovered by Usukhbayar (Usukhbayar and Choi, 2020). Another factor related to pressure is work stress which was proposed by Man et al. (Man *et al.*, 2021).

As presented above, research has revealed a wide spectrum of factors that contribute to a project's safety performance. These form a complex composition of organizational, managerial, legislative, environmental and personnel factors, in addition to project pressure.

2.3 System Dynamics

System dynamics is a method of mathematical modelling whose goal is to enhance decision-making and aid in the understanding of complicated dynamic systems. The method is based on feedback system theory (University of Bergen, 2022).

2.3.1 Systems Thinking: System Dynamics

Systems thinking is the method for understanding large complex systems and their dynamic behavior. It was developed more than 60 years ago by Jay Forrester at the Massachusetts Institute of Technology (Forrester, 2007). Forrester said the benefit of system dynamics was its unique ability to represent the real world (Forrester, 1994). Despite the benefits Forrester also acknowledges the issues that can occur when converting a real-life situation into a simulation.

System dynamics is way of understanding the change and complexity of a system over time (Bala, Arshad and Noh, 2017). The focus is on understanding the entire system and how the different parts interact and affect each other rather than focusing on one part in isolation. System dynamics's methodology is created to handle non-linearity, multiloop and time-lag characteristics in complex dynamic systems. This is done using feedback concepts. Modelling and simulating complex dynamic systems can aid in the understanding of dynamics of systems.

There are two types of systems, open and feedback. Open systems have input that affects output but the output does not affect the input. The system in "unaware" of its performance. In a feedback system the output is used further to produce new output and gouge the performance of the system. Feedback systems can be either positive feedback systems or negative feedback systems.

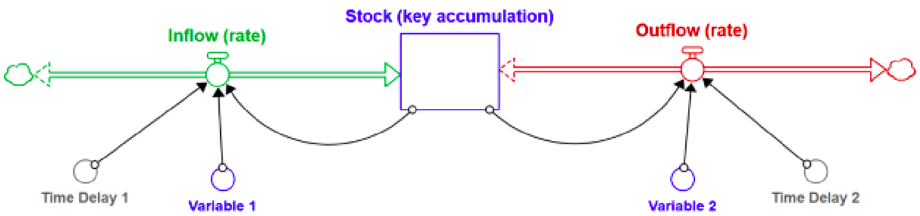


Figure 5: Basic structure of a feedback loop (University of Bergen, 2022)

In figure 5 an example of a simple feedback loop is shown. The value of the stock will be dependent on the inflow, but the inflow is also dependent on the stock value, time and other variables. This loop between the stock and the inflow is a feedback loop as it sends information backwards. The outflow will be determined by the stock as well as time and other variables. Feedback loops are an important part of system dynamics and all systems will contain multiple feedback loops.

A causal loop diagram is a method for visualizing the relationships between variables. A positive causal relationship between two variables indicates that they vary in the same direction, i.e., an increase in one variable causes an increase in the other. This also means a decrease in the first variable will lead to a decrease in the second variable. If there is a negative causal relationship between two variables, then an increase in the first variable will inversely result in a reduction in the second variable. The diagram shows variables with arrows between them. An arrow is marked with either a plus or a minus symbol. This is to indicate if the relationship is positive or negative. The direction of the arrow indicates the direction of the relationship. This type of diagram was first formally used in *The second cybernetics: Deviation-amplifying mutual causal processes* by Magoroh Maruyama in 1963 (Maruyama, 1963). This diagram is an important aid in visualising the connections between variables in systems. An example of a causal loop diagram is shown in figure 6.

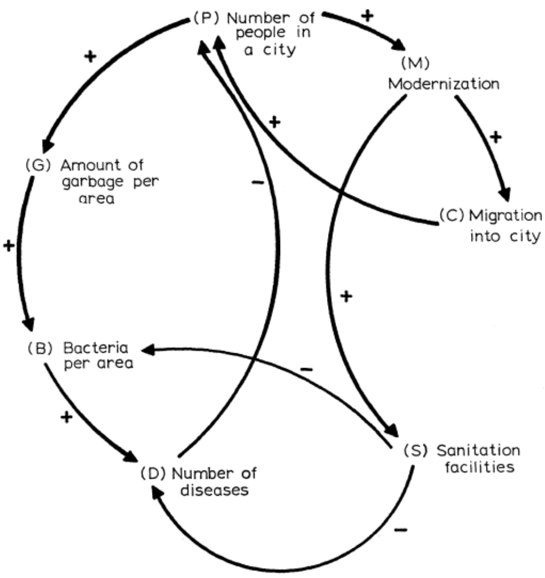


Figure 6: Causal loop diagram (Maruyama, 1963)

System thinking and system dynamics are tools to create models that recreate or simulate behaviors. A model should resemble the reality as closely as is practical, but be reasonably cheap to develop and operate (Fellows and Liu, 2003). It is the structure of the system that generates behavior. Information accessibility, decision-making rules, and other factors may have an impact on the system's dynamics (Sterman, 2013).

2.3.2 Project Simulation

Project simulation is a way of imitating real-world conditions and analyzing project performance outcome or behaviour. Simulations involves some elements of dynamism because it models a process rather than an object (Fellows and Liu, 2003). One example of this is the Virtual Design Team (VDT) model which is a computation model of project organizations (Jin and Levitt, 1996). This was done to identify the needs brought on by various project activities and what additional activities required to be carried out as a result. The critical path method (CPM) can be used to show which activities precede other activities, but CPM is unable to show dependencies of concurring activities and it is also unsuitable for showing dependencies between actors. Therefore there was need for a new model with more complexity.

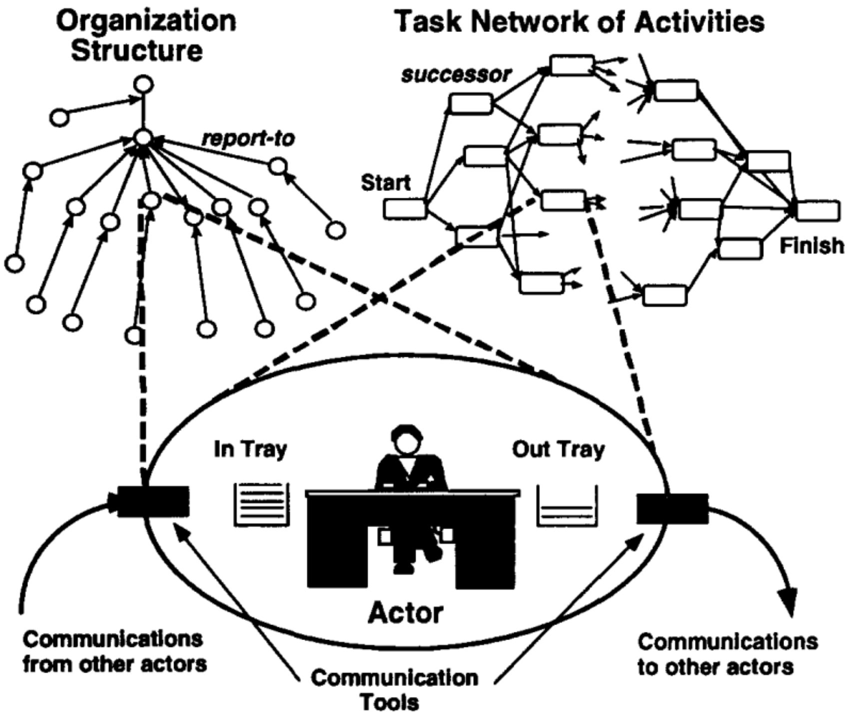


Figure 7: Overview of the VDT model (Jin and Levitt, 1996)

Figure 7 shows the VDT model by Jin and Levitt. The actors, task network, organizational structure and communication flow are the major elements of the information-processing system that is depicted. The model is based upon that an actor has activities on their in tray and on their out tray. The actor based their choice of activity either on time of arrival, by priority or at random and this again caused different needs throughout the organization. The VDT model has organization structure, communication tools, description of project team members and descriptions of activities and their dependencies as inputs. The outputs of the model are project performance, project duration, project cost and project quality.

Other studies have used system dynamics to simulate projects. Depending on what type of project is being simulated, the factors change. For example, construction projects will have some differences from software development projects. In the study *System Dynamics Modeling Strategy for Civil Construction Projects: The Concept of Successive Legislation Periods* by Jing et al. system dynamics was used to simulate construction projects (Jing *et al.*, 2019). The study looked at a project's performance in terms of project cost and project schedule during different legislative periods. The system looked at factors such as the rate of rework, acceptance rate, contract value, contractor adequacy and owner adequacy. In this project simulation the focus was on rework rate and errors found.

Xie et al. created a system dynamics model that simulated a construction project in the study *Effects of Payment Delays at Two Links in Payment Chains on the Progress of Construction*

Projects: System Dynamic Modeling and Simulation (Xie *et al.*, 2019). This model concentrated on cash flow inside a project and payment delays. The study used 66 factors where most of them were related to the project's finances. The factors looked at were fund balance, payment cycle, penalty factor for delay and current penalty amount. The simulation looked at impact of payment delays. The study *Modeling social sustainability in construction projects by integrating system dynamics and fuzzy-DEMATEL method: a case study of highway project* by Rostamnezhad et al. looked at simulating social sustainability in a highway construction project (Rostamnezhad *et al.*, 2020). The model looked at 34 factors divided into five different categories. Each category constitute one subsystem, which are all included in the system dynamics model. The five categories are shown below;

- Stakeholders' engagement factors
- Workforce needs consideration
- Safety related factors
- Health related factors
- Management considerations

System dynamics has also been used to simulate software development projects. The study *System dynamics in software project management: towards the development of a formal integrated framework* looks at creating a system dynamics project-management integrated model (Rodrigues and Williams, 1997). Three system dynamic models are used to model the management process. One for strategic planning and control, one for monitoring, and one for planning. The model illustrates how the operational plan of the management process is impacted by perceived work progress from the engineering process and vice versa. The aim of the model is to see how poor strategic management and related human factor can lead to failures. A different study by Rodrigues and Williams looked at how client behavior affects project performance (Rodrigues and Williams, 1998). Clients can affect the project by imposing schedule restrictions on milestones, high demand on progress reports, delays in approving documents, and changes in work scope through the life cycle. A systems dynamics model was created in an attempt to consider and quantify these effects.

There are many methods for simulating projects. The factors selected for the model will change based on the desired outcome of the model. Some project simulations are generic and not always related to a specific project type. Still, simulating construction projects will often vary from simulating software development projects. Depending on the purpose of the simulation, there will be significant variations even within these project types.

2.3.3 System Dynamics and Safety Performance

System dynamics has been utilized in different studies to simulate the safety performance of construction projects. There are numerous aspects to take into consideration while modeling a construction project. Both project management factors and pressure-producing factors such as delays, rework, and lost time are often taken into account. These system dynamic models frequently consider aspects that are specific to safety, such as hazard awareness, safety training, and risk assessment. What factors are included in the model is often dependent on what effects the study is investigating. Table 1 shows some of the most common system dynamics indicators.

Indicator	Article(s)
Pressure	Mohammadi and Tavakolan, 2019 Qayoom and HW Hadikusumo, 2019 Jiang, Fang and Zhang, 2015 Han *et al.*, 2014 Sun *et al.*, 2019
Cost	Li *et al.*, 2022
Schedule	Li *et al.*, 2022
Safety Climate/Culture	Mohammadi and Tavakolan, 2019 Li *et al.*, 2022 Han *et al.*, 2014
Management Commitment to Safety / Attitude towards safety	Qayoom and HW Hadikusumo, 2019 Li *et al.*, 2022 Han *et al.*, 2014 Mohamed and Chinda, 2011
Supervision	Mohammadi and Tavakolan, 2019 Qayoom and HW Hadikusumo, 2019 Han *et al.*, 2014 Gong *et al.*, 2021 Su *et al.*, 2021 Jiang *et al.*, 2023 Huang *et al.*, 2022
Safety Training	Mohammadi and Tavakolan, 2019 Qayoom and HW Hadikusumo, 2019 Li *et al.*, 2022 Jiang, Fang and Zhang, 2015 Han *et al.*, 2014
Risk Assessment	Qayoom and HW Hadikusumo, 2019
Equipment Condition	Sun *et al.*, 2019
Incident Learning	Mohammadi and Tavakolan, 2019 Jiang, Fang and Zhang, 2015
Worker Competency	Qayoom and HW Hadikusumo, 2019

Table 1: Indicators from the system dynamics and safety performance literature review

Work Pressure

The relationship between pressure and safety performance has been examined in multiple studies. A study by Mohammadi and Tavakolan looked at how production pressure affected the safety performance of a project (Mohammadi and Tavakolan, 2019). Figure 8 illustrates Mohammadi and Tavakolan's system dynamics model. The model demonstrates how elements including labour hours and delays, as well as working speed, human mistake, rework, and lost time, both affect and are affected by production pressure. Rework has a direct impact on the project's incident rate. The model demonstrates how increased production pressure reduces safety through fatigue and rework. The study *Toward an understanding of the impact of production pressure on safety performance in construction operations* identified schedule delays and rework as critical values affecting safety performance (Han *et al.*, 2014). Han et al. looked at how factors such as work pressure and safety climate affected the incident rate. Work pressure would affect production rate and the amount of errors made which could lead to a schedule delay which could lead to more work pressure. Li et al. looked into how helmet use was affected by productivity pressure among many other things (Li *et al.*, 2022). The factor productivity pressure is affected by both cost and schedule in Li et al.'s model. High productivity pressure leads to high workplace stress. Excessive workplace stress is the main factor preventing construction site workers from wearing helmets. High workplace stress also leads to a decline in safety awareness which also leads to nonhelmet use behavior (Li *et al.*, 2022).

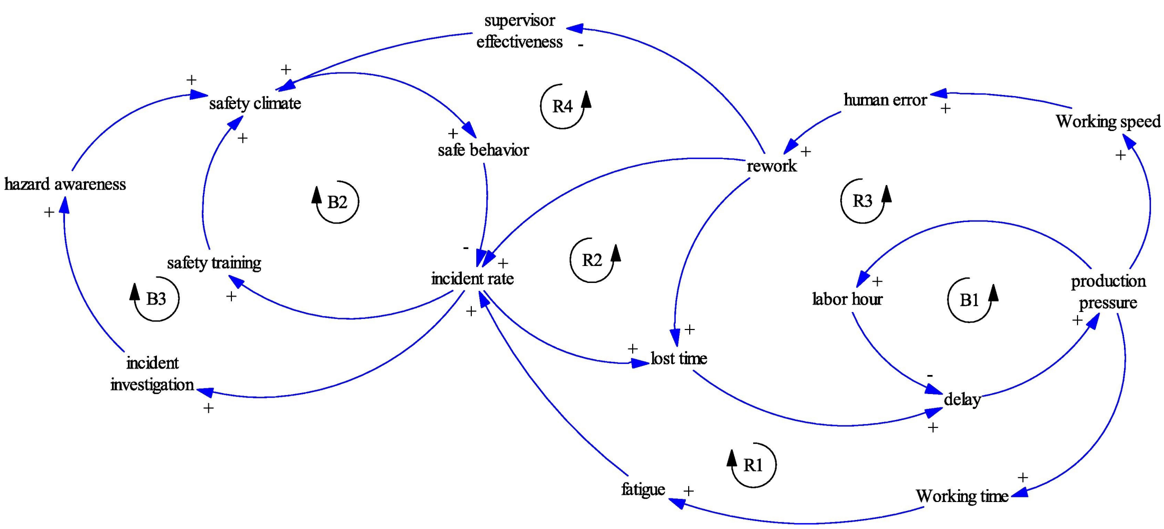


Figure 8: System dynamics model (Mohammadi and Tavakolan, 2019)

Safety Climate and Safety Culture

Many studies have looked at how safety climate or safety culture affects safety performance. The study *Multilevel safety culture affecting organization safety performance: a system dynamic approach* looked at safety level on three different levels; strategic, tactical and operational. Combined, these three levels create a multilevel safety culture (Qayoom and HW Hadikusumo, 2019). The model assumes that the company's dedication to safety will be passed down from top management to middle management and finally to the operational level. Each level is represented with its own sub-model which are all combined in one system dynamics model. The top management safety culture affects the middle management safety culture which again affects operational-level safety culture. This can be seen in figure 9. The middle management safety culture affects risk levels which affects unsafe conditions. The operational-level safety culture leads to unsafe acts. The combination of unsafe acts and unsafe conditions determine the incident rate (Qayoom and HW Hadikusumo, 2019).

The three sub-models contains factors that determine the safety culture for each level of management. The top management safety culture is centered around finding the balance between productivity and safety priority. This part of the model contains factors such as focus on safety, safety policy and risk assessment. Safety policies, goals and targets are set by the top management, then responsibility is shifted to the middle management. The middle managements safety culture is dependent on safety related factors such as safety training, safety supervision and hazard identification. At this level it is the safety manager's responsibility to take actions such as safety training to reduce risk levels. The operational level safety culture is where factors such as safety compliance are located. Factors for determining the safety culture at this level is worker competence, worker experience and safety supervisor competence (Qayoom and HW Hadikusumo, 2019).

Figure 9: System dynamics model (Qayoom and HW Hadikusumo, 2019)

Project Management

Numerous other studies have also looked at the connection between project management and incident rates. The study *Understanding the Causation of Construction Workers' Unsafe Behaviors Based on System Dynamics Modeling* examined what drives workers to commit unsafe behaviors (Jiang, Fang and Zhang, 2015). Individual conditions, environmental conditions and management conditions were looked at. Management conditions include safety communication, safety resources, incident learning and safety training. Some of the individual conditions were safety awareness and safety knowledge. These models illustrate that bad safety management leads to unsafe behaviour from workers which again leads to an increase in incidents. Additionally, the study *System dynamics modelling of construction safety culture* investigated five key enablers of construction safety culture (Mohamed and Chinda, 2011). It was discovered that in order to improve safety culture in the construction industry the primary focus should be on enhancing leadership attributes.

Supervision

Supervision is an important tool for the management to contribute to a better safety performance. Gong et al. looked at how the probability of government supervision was affecting the probability of contractors making sufficient safety investments (Gong *et al.*, 2021). Su et al. created a model which investigates a construction safety standard system and what is being affected if such a system was made (Su *et al.*, 2021). It also looked at the probability of obeying construction safety standards (CSS). Important factors were the probability of supervision, the contractor's payoff for not obeying CSS, supervision accuracy and cost of safety supervision. Huang et al. examined how several factors affected both the unsafe behaviour probability and the positive supervision probability in construction projects (Huang *et al.*, 2022). The study *Using evolutionary game theory to study construction safety supervisory mechanism in China* by Jiang et al. looked at supervision on Chinese construction projects (Jiang *et al.*, 2023). The model looked at the rate of inspection from a government supervision agency, the rate of supervision from the supervision engineer and the safety investments from the general contractor and how this affected the probability of accidents.

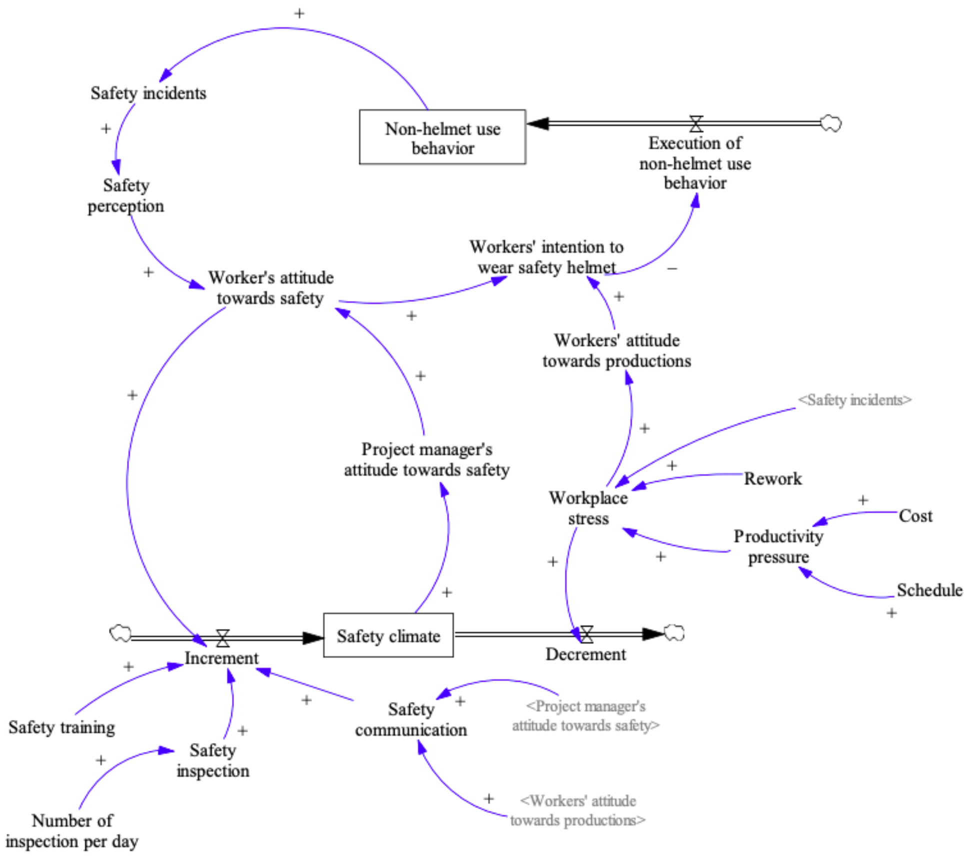


Figure 10: System dynamics model (Li *et al.*, 2022)

Safety Related Indicators

When simulating the safety performance of a project it is often necessary to include factors directly related to project safety. The study *Understanding the influence of safety climate and productivity pressure on non-helmet use behavior at construction sites: a case study* by Li et al. looked at safety training, number of inspections per day, safety communication and the effects this had on the safety climate (Li *et al.*, 2022). Their model can be seen in figure 10. Sun et al. looked at factors such as management influences, pressure accumulation, skills training and equipment condition (Sun *et al.*, 2019). This study also investigated accident prevention through design and therefore has additional factors to simulate this. These factors include access limitations, direct constraints on response, direct constraints on objects and structural boundary of an object imposed by its environment.

Calibrating and Validating

The system dynamics models were validated in various ways. In some of the studies the simulations were compared to collected data (Li *et al.*, 2022). In other studies the simulations went through tests such as sensitivity testing. This means varying the model's input parameters and tracking how the output changes as a result. In the study by Qayoom and HW Hadikusumo the simulation went through an extreme condition test. This was to ensure that the "model should be realistic regardless of the extreme inputs" (Qayoom and HW Hadikusumo, 2019). They tested if the model's output was still realistic after providing it with extreme input values. Mohammadi and Tavakolan used behavior reproduction and a sensitivity analysis to validate their model. They compared the simulated results with actual accident data to compare if the simulation produced similar results (Mohammadi and Tavakolan, 2019).

2.4 AI in Projects

AI is a field of computer science that involves simulating human intelligence processes using machines. There is no single inventor of AI, the field emerged as a result of the efforts of many researchers and scientists over several decades. The first mathematical model for a neural network was published by Walter Pitts and Warren McCulloch in 1943 (McCulloch and Pitts, 1943). This is commonly considered to be the beginning of AI (Mohammed, Khan and Bashier, 2016). This model demonstrated how a neural network composed of simple elements could have enormous processing capacity. This model drew from Alan Turings theory of computation (Russell and Norvig, 2016). Turing was a mathematician and logician and he is one of the earliest pioneers in AI. In 1950 he asked the question "Can machines think?" in his paper *Computing machinery and intelligence* (Turing, 2009). He introduced the concept of testing the intelligence of a machine (Mohammed, Khan and Bashier, 2016). This intelligence test for machines was called the Turing test (Russell and Norvig, 2016). John McCarthy, Marvin Minsky, Nathaniel Rochester and Claude Shannon organized the first summer AI conference in 1956, which was held at Dartmouth College

(Mohammed, Khan and Bashier, 2016). The conference brought together a group of researchers who were interested in exploring the potential of artificial intelligence. The actual term artificial intelligence was first coined at this conference by John McCarthy (Smith *et al.*, 2006). Since 1956, advances in AI have been made consistently. This combination of improved computational power and developments in AI means the possibilities of AI are expanding.

AI has many different uses. One example is utilizing AI in projects. To be able to use AI in projects a challenge is often the collection of data. Good quality data needs to be collected or else it is unsuitable for AI. The study *Use of big data in project evaluations* by Olsson and Bull-Berg looked at how different data from projects could be used for evaluations. They state that despite the volume of data generally increasing, access to relevant data could still be a challenge. It is the combination of availability and quality that makes data suitable for evaluating projects (Olsson and Bull-Berg, 2015). Another challenge when implementing AI is that organizations don't have a necessary readiness level to effectively implement AI (Alsheibani, Cheung and Messom, 2018). A study by Wang looked at why people are sceptical about using AI and identified three main reasons: people are afraid that AI will take away jobs, people fear AI will cause failures and people fear the abuse of AI (Wang, 2019). These factors all limit the use of AI.

2.5 Machine Learning

Machine learning is the ability of computers to learn from data (Wade and Glynn, 2020). There are three types of learning; supervised, unsupervised and reinforcement learning (Russell and Norvig, 2016). Supervised learning describes when the agent receives input-output pairs and learns a function to map from input to output. Unsupervised learning is when the agent learns patterns from the input data without getting any feedback, such as clustering. Reinforcement learning is when the agent learns from punishment and rewards. Within supervised learning there are two different learning problems; classification and regression. Classification describes the problems where the input is assigned into predefined categories, while in regression the output is a continuous value (Jo, 2021).

2.5.1 Synthetic Data

Supervised learning can be performed on both real and synthetic data. Synthetic data is "generated by a computer program rather than being collected from real-world sources" (Birisci, Gursakal and Celik, 2023). Edali and Yücel state that machine learning can be used to make inferences from raw data, which can be gathered through any data generation or collection from sensors, consumer transactions or simulation models (Edali and Yücel, 2020).

Japkowicz and Shah argue how artificial data can be an advantage to gather insights of the behaviour of the algorithms. Even though real data provide valuable information about the real world, artificial data has other advantages. It allows for exploration of variabilities which are realistic but has no available data (Japkowicz and Shah, 2011). Synthetic data can also be simulated in a controlled manner for specific purposes. This provides more control when conducting experiments (Japkowicz and Shah, 2011). Nevertheless, they highlight how artificial data run the risk of oversimplifying problems.

Nikolenko describes how many of the problems in AI are caused by insufficient data. Often, the datasets are too small or the data requires manual labelling, which is very time expensive. A solution to this is generating synthetic data. An advantage of synthetic data is that it can be produced during training, which allows for not having to store huge datasets over time (Nikolenko, 2021). A common approach is training the model on synthetic data with the intention of using it on real data (Nikolenko, 2021).

Synthetic data can also be used to increase the performance of an algorithm. An example is resampling the training set with artificial data. Here, the artificial data is often a manipulation of real training samples. This can balance the distribution of the categories and help remove the bias towards a certain category in classification problems (Jo, 2021). The artificial data is called *virtual training examples*. The process of modifying real data is often referred to as *data augmentation* (Nikolenko, 2021).

At last, synthetic data is suitable for solving privacy or legal issues (Nikolenko, 2021). In some fields it can be difficult to use real data, such as within healthcare, finance and other social sciences. Synthetic data is also appropriate when dealing with privacy guarantees.

2.5.2 Machine Learning Models

There is a wide spectrum of machine learning algorithms to be used to create a model. A model is an algorithm instantiated with data and parameters (Zhou, 2021). Common machine learning algorithms are decision trees (DT), random forest (RF), AdaBoost, XGBoost and support vector machines (SVM). These are described below.

Decision Trees

The decision tree algorithm is a supervised machine learning algorithm (Jo, 2021). The decision tree is a function that returns a single output from a list of attribute values (Russell and Norvig, 2016). The algorithm creates a model with a tree structure with internal nodes and leaf nodes. Each internal node serves as a test for an attribute within the function. Each branch becomes an output of the test and each leaf node represents a class label. The decision tree is built using the training set and classifies items following its branches from the root node to a leaf node (Jo, 2021). The class label of this leaf node is the prediction. Figure 11 shows an example of a decision tree used for deciding whether or not to wait for a table at a restaurant.

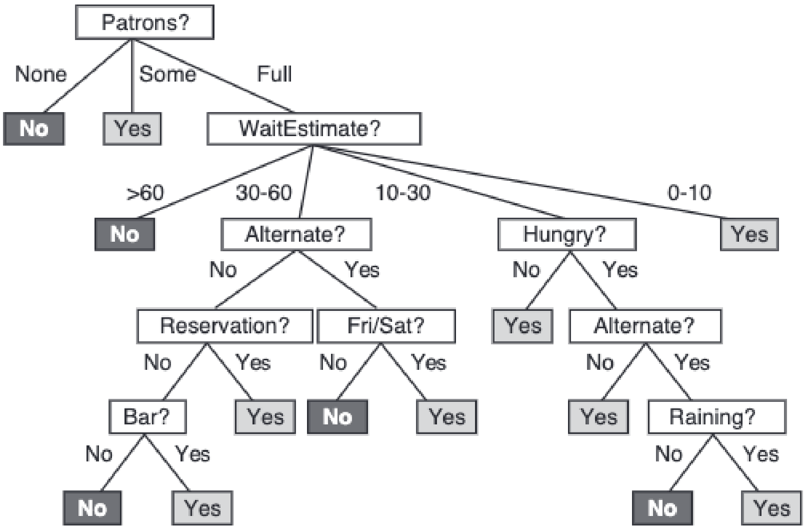


Figure 11: A decision tree for deciding whether to wait for a table (Russell and Norvig, 2016)

As mentioned a decision tree is developed using a test set. Choosing the right attributes is important in order to get the highest performing decision tree possible. The algorithms employ a greedy divide-and-conquer strategy in which the attributes deemed to be most significant are tested first and the remaining sub-problems are then addressed (Russell and Norvig, 2016). For these sub-problem the most important attribute is again selected and used for testing the subset. This is how the decision tree is created. It is therefore crucial to select important attributes that create a balanced tree. The importance of the attributes is measured using information gain. The attribute with the highest information gain is selected as the root node. Entropy is used in information gain calculations. Entropy measures the uncertainty of a variable and varies from zero to one where zero means no uncertainty (Russell and Norvig, 2016). Equation 1 shows that information gain is calculated by subtracting the attributes expected remaining entropy after the split from the current entropy previous to the split. Equation 2 shows the calculations for entropy, where V is a random attribute with values v_k and with a probability of $P(v_k)$ for each value.

$$Gain(A) = Entropy - Remainder(A) \tag{1}$$

$$Entropy = H(V) = \sum_k P(v_k) \log_2 \frac{1}{P(v_k)} = -\sum_k P(v_k) \log_2 P(v_k) \tag{2}$$

Equation 3 uses a boolean variable that is true with probability q to define entropy. A split point is determined for continuous variables. This equation can be adjusted to decision trees and that gives equation 4, with p the number of positive values and n the number of negative values. This calculates the entropy of an output variable for the entire set. After testing attribute A the remainder of the expected entropy is calculated using equation 5.

$$B(q) = -(q \log_2 q + (1-q)\log_2(1-q)) \tag{3}$$

$$H(output) = B(\frac{p}{p+n}) \tag{4}$$

$$Remainder(A) = \sum_{k=1}^{d} \frac{p_k + n_k}{p+n} B(\frac{p_k}{p_k + n_k}) \tag{5}$$

The decision tree model is a tree structured model created by the decision tree algorithm. Decision trees provide many advantages, including robustness and intuitiveness (Jo, 2021). Still, the greedy search results in the model does not always obtain the optimal accuracy. Furthermore adding one more instance can change the structure of the entire decision tree (Russell and Norvig, 2016).

Random Forest

The random forest model is an *ensemble* machine learning model. An ensemble method combines the predictions from a collection of hypotheses and creates one prediction based on this collection (Russell and Norvig, 2016). The ensemble method used in the random forest model is *bagging*. The training set is randomly split up into subsets and for each subset a decision tree is created. The random forest is the collection of these decision trees (Jo, 2021). Since the training set is split up many of the trees will be trained in a weaker way (Bonaccorso, 2017). Figure 12 shows how the training set is split into multiple random subsets that are then built into different decision trees.

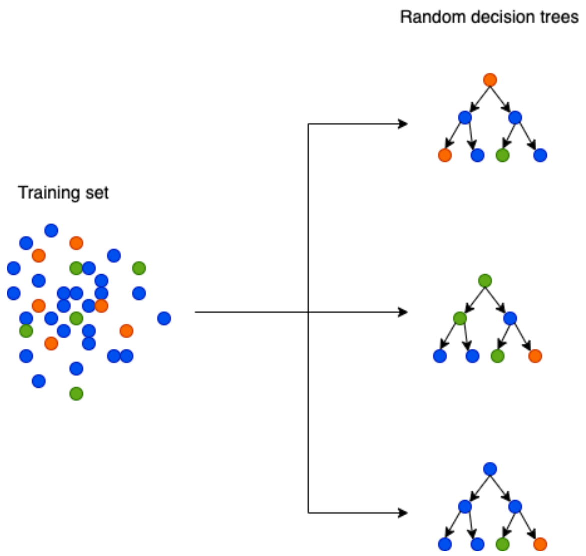


Figure 12: Splitting up the training set (Borkenhagen and Olsen, 2022)

Each decision tree makes a prediction in order to predict an instance of the test data. These predictions are done in parallel. The predictions from the subsets are then combined into an aggregated result. The most frequent class is then selected for classification. This selection is called voting. For regression the average of all the predictions is calculated (Jo, 2021). Figure 13 shows how the different predictions are combined into one result. This reduces the risk of overfitting in comparison to the decision tree model, since the predictions are based on multiple decision tree models which reduces the variance (Bonaccorso, 2017). This aggregated result is the result of the random forest model.

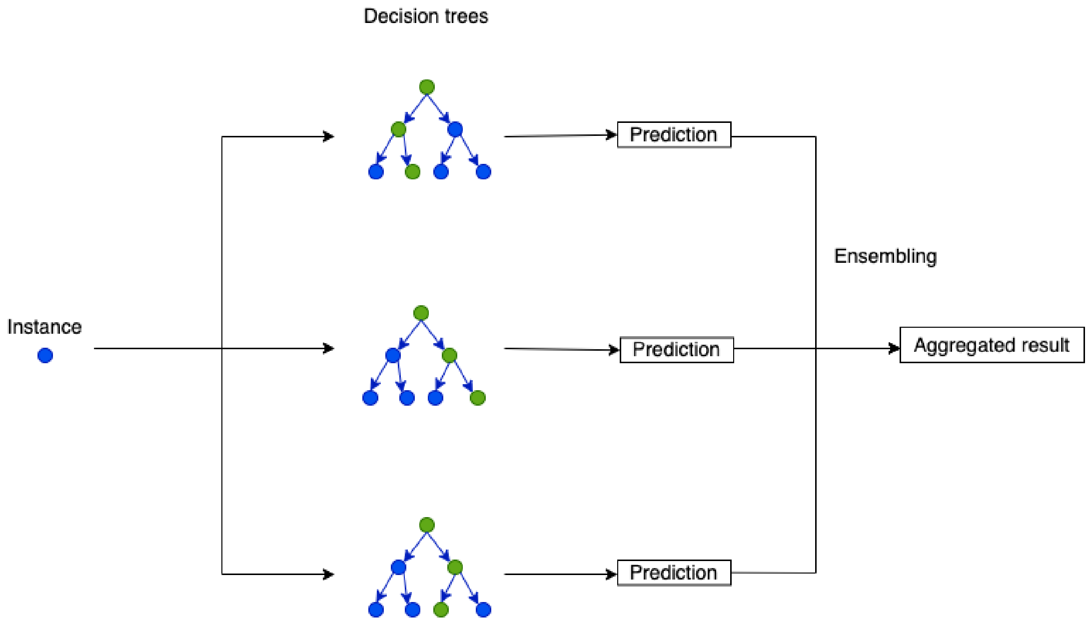


Figure 13: Combining the predictions (Borkenhagen and Olsen, 2022)

AdaBoost

AdaBoost is an algorithm that utilizes adaptive boosting. *Boosting* is the process of converting weak learners to strong learners (Zhou, 2021). This is done by first training a base learner and then adjusting the distribution of the training samples according to the results of the hypotheses (Bonaccorso, 2017). This way, the incorrect samples will be weighted more than the correctly classified samples and will therefore gain more attention in the next hypotheses (Wade and Glynn, 2020). This weighting process will continue until the algorithm classifies the remaining samples correctly or until a predefined value. At last, the hypotheses are weighted based on their performance and combined in an ensemble for the final prediction. The boosting process is shown in figure 14.

AdaBoost is a boosting algorithm with decision stumps as its hypotheses. Decision stumps are decision trees containing one root node with two leaves. The training process goes as described above, with an iterative process of creating new decision stumps. These are based on the weighting of the samples. Finally, the decision stumps are weighted based on their performance and combined in order to make the final predictions. Equation 6 shows the linear combination of the base learners, with x being the input, h_i the hypotheses and z_i the weights of each hypothesis (Russell and Norvig, 2016). *h(x)* is the final prediction.

$$h(x) = \sum_{i=1}^{K} z_i h_i(x) \tag{6}$$

One of AdaBoost's qualities is it's ability to transform a weak learner into a perfect classifier (Russell and Norvig, 2016). This happens when the original algorithm is a weak learner, meaning it performs better than random guessing. The result will then be a classifier which classifies the entire training set correctly.

Figure 14: Boosting (Borkenhagen and Olsen, 2022)

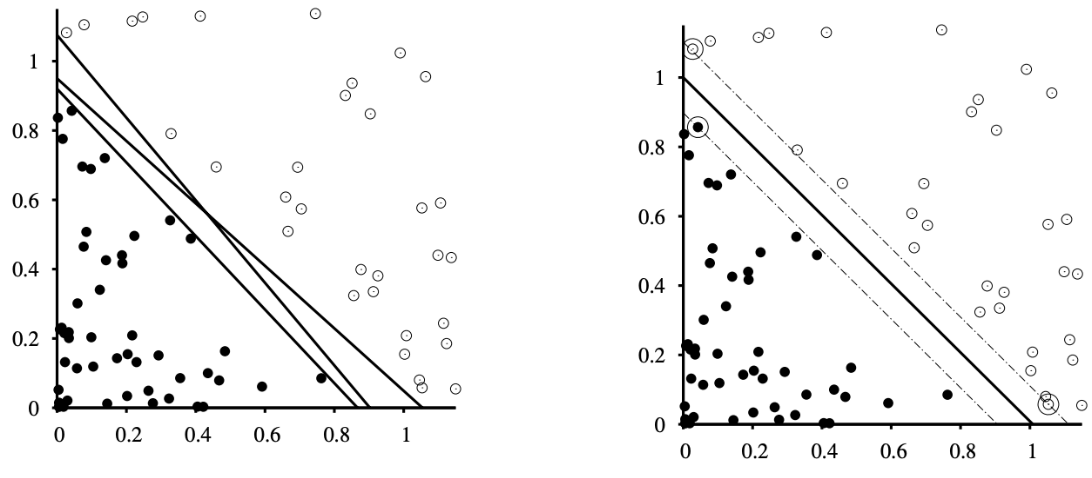
XGBoost

Another boosting algorithm is XGBoost, standing for eXtreme Gradient Boosting (XGB). XGBoost incorporates the structure of *gradient boosting* (Wade and Glynn, 2020). Gradient boosting allows for building a tree ensemble such that a target loss function is minimized (Bonaccorso, 2017). It uses gradient descent to develop new hypotheses. This means each new tree is focusing on the error of the previous trees. Then it sums the residuals of each tree, the differences between the predicted and actual values, to score the model (Wade and Glynn, 2020). Similarly to AdaBoost, gradient boosting transforms weak learners to strong learners. It does however differ from AdaBoost by focusing on a global goal instead of re-weighting the data samples.

XGBoost is an advanced version of gradient boosting. It is commonly used as it is faster and delivers better results than traditional gradient boosting (Wade and Glynn, 2020). XGBoost builds the tree ensemble by iteratively building decision trees and moving towards the gradient that minimizes the defined loss function (Russell and Norvig, 2016). The algorithm utilizes regularization, pruning and parallel computing. Regularization is used to avoid overfitting by penalizing complex hypotheses to give room for the more simple ones. Pruning is the act of removing certain possibilities without having to examine them (Russell and Norvig, 2016). Parallel computing occurs when several units are working together on the same problem at the same time (Wade and Glynn, 2020). XGBoost is particularly popular for factored tabular data (Russell and Norvig, 2016).

SVM

The support vector machine (SVM) algorithm attempts to create a hyperplane to classify the data points (Russell and Norvig, 2016). The algorithm aims to minimize the expected generalization loss. This is accomplished by selecting the separator that is furthest away from the samples seen so far. Three candidate separators are shown in figure 15a. The selected separator is called the *maximum margin separator*, and shown as the heavy line in figure 15b. The maximum margin separator serves as the decision boundary that separates the different classes. The circles in the figures represent the data points, with their classes distinguished by their color. Each sample will be classified based on their location in regards to the maximum margin separator. The samples that are closest to the separator are called support vectors (Zhou, 2021). The area between the dotted lines is called the margin. The maximum margin separator is the one that has the largest margin (Russell and Norvig, 2016).



(a) Three candidate linear separators

(b) The maximum margin separator

Figure 15: SVM (Russell and Norvig, 2016)

The SVM algorithm finds a linear hyperplane, that being the maximum margin separator. When working with non-linear problems, the algorithm uses the *kernel trick*. By projecting the data points into a higher dimensional space, the kernel trick allows for separation of data points which are not necessarily linearly separable in the original input space (Zhou, 2021). This is an important property of SVM, as it is then able to solve non-linear problems as well (Bonaccorso, 2017). In addition, the SVM algorithm is non-parametric, meaning it preserves training samples. However, it only preserves a small fraction of them, meaning it gains the advantages of both non-parametric and parametric models (Russell and Norvig, 2016).

2.5.3 Model Evaluation Metrics

There are many ways to evaluate classification performance. This section will describe confusion matrices and types of classification predictions with an emphasis on binary classification. Further, the metrics accuracy, precision, recall and F1-score will be presented.

Confusion Matrix

A confusion matrix can be used to visualize the test-set performance of a classifier. It is an l x l matrix for a dataset with l classes. The confusion matrix can be denoted $\mathcal{C}$, while the classifier *f*. Each element c_{ij} in the matrix denotes how many examples of the label i were predicted as label j. Each test sample is defined as $\mathbf{x}$, while y is its corresponding label. The confusion matrix is defined in equation 7 (Japkowicz and Shah, 2011);

$$\mathbf{C}(f) = \left\{ c_{ij}(f) = \sum_{\mathbf{x} \in T} [(y = i) \wedge (f(\mathbf{x}) = j)] \right\} \tag{7}$$

From this equation, $\sum_{j=1}^{l} c_{ij}(f) = c_{i.}(f)$ is the total number of samples of class i in the test set. Further, $\sum_{i=1}^{l} c_{ij}(f) = c_{.j}(f)$ is the total number of samples labeled correctly to class j. All the diagonal entries c_{ii} are the correctly classified samples for class i. This means all nondiagonal entries are misclassified.

Table 2 shows this distribution for binary classification, where there are l=2 number of classes and *Act* stands for actual and *Pred* stands for predicted.

f	Pred_Negative	Pred_Positive
Act_Negative	$c_{11}(f)$	$c_{12}(f)$
Act_Positive	$c_{21}(f)$	$c_{22}(f)$

Table 2: Confusion matrix for binary classification (Japkowicz and Shah, 2011)

Binary Classification

In the binary classification case, all predictions can be classified as one of the following; true negative (TN), false negative (FN), true positive (TP) or false positive (FP) (Zhou, 2021). Redefining table 2 with respect to this gives the following matrix (table 3);

f	Pred_Negative	Pred_Positive
Act_Negative	TN	FP
Act_Positive	FN	TP

Table 3: Confusion matrix for binary classification with prediction labels

With labels being negative (0) and positive (1), this can be translated into what is shown in table 4. In other words, a true positive is when the sample is positive and is predicted accurately. A true negative is when the sample is negative and predicted accurately. A false positive is when the sample is predicted as positive, but predicted incorrectly. At last, a false negative is when the sample is predicted as negative, but predicted incorrectly.

Predicted label	Actual label	Definition
1	1	True Positive
0	0	True Negative
1	0	False Positive
0	1	False Negative

Table 4: Binary classification predictions

Japkowicz and Shah state that the most natural way to measure a learning algorithm in the single class scenario is measuring the true positives. This is also called the *sensitivity* of the classifier (Japkowicz and Shah, 2011). This term comes from the healthcare industry, as one would typically be interested in how many of the patients that has a disease one would be able to detect. The compliment of this, measuring the true negatives, is referred to as *specificity*. In the health context, this would be how many of the healthy patients that tests negative. The formulas for sensitivity and specificity are defined below (equation 8, equation 9);

$$\text{Sensitivity} \;=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{8}$$

$$\text{Specificity} \;=\frac{\text{TN}}{\text{FP}+\text{TN}} \tag{9}$$

Accuracy

Accuracy is the most frequently used evaluation metric for classification performance (Japkowicz and Shah, 2011; Hossin and Sulaiman, 2015). It measures the fraction of correctly classified instances. In terms of the entries in the confusion matrix, accuracy can be defines as (equation 10);

$$\text{Acc}(f)=\frac{\sum_{i=1}^{l} c_{ii}(f)}{\sum_{j=1}^{l} c_{ij}(f)} \tag{10}$$

Applying this to the binary classification case gives (equation 11);

$$\text{Acc}(f)=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{11}$$

Accuracy is an effective metric in regards to overall performance, but does however have some shortcomings. It does not take into occurrence that prediction of different classes might be of different importance and the dataset being imbalanced.

Precision

Precision, also called the positive predictive value (PPV), measures how many of the predicted positives that are accurate. Precision is also an important measure in the health context (Japkowicz and Shah, 2011). The precision on a given class of interest i can be defined as (equation 12);

$$\mathrm{PPV}_i(f) = \mathrm{Prec}_i(f) = \frac{c_{ii}(f)}{\sum_{j=1}^{l} c_{ji}(f)} = \frac{c_{ii}(f)}{c_i(f)} \tag{12}$$

When precision is applied to binary classification it can be defined as (equation 13);

$$\mathrm{Prec}(f) = \mathrm{PPV}(f) = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{13}$$

Hence, precision is the ratio between the true positives and all positive predictions.

Recall

Recall is another name for sensitivity, usually referred to in the context of information retrieval. Hence, recall measures the fraction of actual positives that are correctly classified (Hossin and Sulaiman, 2015). The equation for recall is shown below (equation 14), with TPR being the true positive rate;

$$\mathrm{Rec}(f) = \mathrm{TPR}(f) = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{14}$$

F1-Score

The F measure represents the harmonic mean between recall and precision values (Hossin and Sulaiman, 2015). There are several variations to the F measure, with one being the F1 score. The F1 score is a balanced measure that weights recall and precision equally. The F1 score equation is shown below (equation 15);

$$\mathrm{F1} = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * \mathrm{TP}}{2 * \mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{15}$$

To summarize, confusion matrices as well as positive and negative prediction outcomes have been presented. The metrics accuracy, precision, recall and F1-score have also been introduced. Accuracy has limitations regarding imbalanced datasets and difference in class importance. Precision and recall on the other hand focuses on only the class of interest, being the positive class. This also implies some limitations, as the true negatives are not taken into account. The same goes for F measures, as they are combinations of precision and recall. Additionally, precision and recall are contradictory, meaning if the recall is high the precision is often low and opposite (Zhou, 2021).

As no single metric is enough to encapsulate all relevant aspects of a classifiers performance, these metrics should be combined (Japkowicz and Shah, 2011). However, utilization of many metrics makes it more difficult to interpret the results.

2.5.4 Machine Learning for Safety Predictions

There have been several attempts to utilize machine learning for safety predictions. This section covers previous objectives as well as previous machine learning models that have been used for this purpose.

Prior Research Objectives

A common objective is predicting accidents before they occur. None of the reviewed articles have researched whether one can predict safety performance based on only planning data. There has however been research done on predicting safety performance based on data from both the planning and the construction phase. Poh, Ubeynarayana and Goh combined data from both phases in their study (Poh, Ubeynarayana and Goh, 2018). The data was a combination of both safety and project management factors. A feature selection method was used to identify the 13 most significant factors that impacts safety. Further, both the number of accidents and their severity was predicted using machine learning. An accuracy of about 78% was obtained. Jafari et al. also performed machine learning on a dataset from the planning and construction phase, but focused on project data only (Jafari *et al.*, 2019). They performed binary classification on whether a project would have an accident or not. An accuracy of 93.5% was obtained.

Andreassen et al. proposed to use the number of HSE incidents that has already happened in a project as a leading risk indicator (Andreassen *et al.*, 2020). Koc, Ekmekcioğlu and Gurgun used time series for prediction of accidents (Koc, Ekmekcioğlu and Gurgun, 2022). Three different time intervals were used; one day, seven days and 30 days. Their research uncovered that the most influential factor was the number of accidents the two previous days, which is in compliance with Andreassen et al.'s proposal. Gregoriades and Mouskos wanted to estimate an accident rate every 15 minutes in a traffic simulation (Gregoriades and Mouskos, 2013).

The prediction of accidents are not always with a forecasting perspective. Gao, Lu and Ren predicted whether an accident had occurred or not on a railway crossing (Gao, Lu and Ren, 2021). This binary classification was done on a dataset containing information until and including the time of the potential accident, and can therefore not be classified as a leading indicator. Ajayi et al. predicted accidents and injury frequency based on project data including employee information, day of the week, time of the incident and type of task (Ajayi *et al.*, 2020). Similarly to Gao, Lu and Ren, this dataset also contained information including the time of the potential accident.

The second, and most widely used perspective is predicting the outcome of an accident. This way, more data is available as the prediction takes place after the accident has already occurred. A common objective is predicting the severity of an accident. For this purpose, data can include the time, location and accident type. Recal and Demirel performed both binary and multi-class classification to predict accident severity (Recal and Demirel, 2021). The binary classification covered whether an accident was fatal or not, while the multi-class divided severity into three classes; minor, major and fatal. Choi et al. also wanted to classify which workers might face a fatality risk based on accident data (Choi *et al.*, 2020). Here, the data included type of construction, date of the accident and information about the worker. Several other studies were conducted to predict injury severity (Sarkar *et al.*, 2019; Sarkar *et al.*, 2020; Gangadhari, Khanzode and Murthy, 2022; Kim and Lim, 2022; Zhang *et al.*, 2022). Further, Lee et al. used machine learning to explore correlations between factors and accident severity (Lee *et al.*, 2020). Zhu et al. also studied factors together with safety performance in a resilience perspective (Zhu *et al.*, 2020).

Other outcomes have also been predicted, such as number of days away, disability status and the type of accident. Yedla, Kakhki and Jannesari wanted to predict both the outcome of an accident together with the predicted number of days away (Yedla, Kakhki and Jannesari, 2020). Kim et al. predicted the number of days away due to an accident and focused on falling accidents only (Kim *et al.*, 2022). Kang, Koo and Ryu conducted a feature importance study with the target being number of lost workdays (Kang, Koo and Ryu, 2022). Koc, Ekmekcioğlu and Gurgun attempted to predict the post-accident disability status (Koc, Ekmekcioğlu and Gurgun, 2021). Another outcome study done by Kang and Ryu was predicting the accident type (Kang and Ryu, 2019). This was also done by Shin (Shin, 2019). Further, Nayak et al. predicted injury type as well as casualty class and occupation (Nayak *et al.*, 2022).

A third and less common objective is predicting the phase of an accident. Jiang et al. predicted tower crane accident phase based on accident data (Jiang *et al.*, 2021). The purpose of predicting the accident phase is to gather more information such that future accidents can be predicted more accurately. Overall, the literature review uncovered that there are many different objectives of applying machine learning for safety predictions. However most research have focused on predicting the outcomes of accidents based on data until and including the accident time.

Prior Research Machine Learning Models

A big amount of machine learning models have been tested through the research presented above. Table 5 shows which algorithms were utilized for this. The bold check marks show which algorithm that gave the best result for a certain article. If only one algorithm was tested, it is not marked in bold. Further, if the machine learning model with the best performance was a combination of two algorithms, neither of the algorithms have been marked in bold. At last, if the article had several top performing algorithms, they were both marked in bold.

		Machine learning models												
Article No.	Author(s)	RF	DT	LR	AB	XGB	CNN	ANN	KNN	BN	MARS	SVM	SGB	GBM
1	Koc, Ekmekcioğlu and Gurgun (2022)							✔			✔	✓		
2	Choi et al. (2020)	✔	✓	✓	✓									
3	Kim and Lim (2022)	✓		✓		✔		✓				✓		
4	Zhu et al. (2020)	✓	✓	✔				✓	✓	✔		✓		
5	Koc, Ekmekcioğlu and Gurgun (2021)	✓			✓	✔								
6	Sarkar et al. (2020)	✔						✓	✓	✓		✓		
7	Lee et al. (2020)		✓									✓		
8	Recal and Demirel (2020)											✓	✔	
9	Kang and Ryu (2019)	✓												
10	Gao, Lu and Ren (2021)	✓	✓				✔							
11	Sarkar et al. (2019)		✓											
12	Ajayi et al. (2020)	✓	✓											✔
13	Kim et al. (2022)							✓						
14	Yedla, Kakhki and Jannesari (2020)	✔	✓	✓				✔						
15	Nayak et al. (2022)									✓				
16	Shin (2019)		✓					✓					✔	
17	Jiang et al. (2021)	✓												
18	Jafari et al. (2019)	✓		✓					✔	✔		✓		
19	Kang, Koo and Ryu (2022)	✓												
20	Gregoriades and Mouskos (2013)									✓				
21	Zhang et al. (2022)	✓		✓		✔			✓	✓				
22	Gangadhari, Khanzode and Murthy (2022)	✓				✔		✓	✓	✓		✓		

Table 5: Machine learning models used in safety prediction articles (Borkenhagen and Olsen, 2022)

The most widely used algorithms were random forest, decision trees, artificial neural networks (ANN) and SVM. Random forest was used in 14 out of the 22 articles. It outperformed the others in three of the articles. Further, the decision tree algorithm was utilized in eight articles. However, it did not perform the best in any of the articles. ANN was also used in eight articles, whereas it had the best performance in two of them. Similarly to decision trees, SVM was widely used, but never a best performing algorithm. An algorithm that stood out was perhaps XGB which was the best performing algorithm in all the studies it had been tested on. It was however not as widely used as the random forest, decision trees, ANN and SVM.

2.6 System Dynamics and Machine Learning

There have been several attempts on combining system dynamics models with machine learning. It has been explored how machine learning can complement traditional modelling to improve system dynamics models (Chen, Tu and Jeng, 2011; Abdelbari and Shafi, 2017; Duggan, 2020; Edali and Yücel, 2020; Weng *et al.*, 2022). Research have also focused on how machine learning can be used on data produced by system dynamics models (Prabhakaran and Martin Jr, 2020; Chen *et al.*, 2022; Roozkhosh, Pooya and Agarwal, 2022).

In 2016, Barlas stated that system dynamics models can also be handled using a wider set of approaches, including machine learning (Duggan, 2020). Edali and Yücel developed a simulation model for influenza epidemics, and explored the model further using machine learning (Edali and Yücel, 2020). Their motivation for using this combination was a more extensive exploration of the model to provide more insights on behavioral aspects. They developed a procedure consisting of three steps;

1. Use random forest to capture input-output relationships on data obtained from the simulation model
2. Improve the random forest metamodel further by active learning and additional simulation runs
3. Extract information from the metamodel to form a set of IF-THEN rules

When this procedure was conducted on the epidemic data, new relationships were uncovered and the results were improved (Edali and Yücel, 2020). Hence, they concluded that their approach could give a better understanding of the model.

Duggan also proposed to apply a combination of machine learning and system dynamics to different phases of the model building. An example proposed was applying these methods for infectious disease outbreaks using Markov Chain Monte Carlo simulation and machine learning approaches to analyze model output (Duggan, 2020). However, he stated that a possible barrier for these applications could be the complexity of machine learning and another focus within the area. Still, he argues how machine learning could be useful to simulation data and presents a simple example using the R language. The example consists of the development of a simple simulation model connected to weather. After a model is developed and a dataset has been simulated, a decision tree algorithm is used to explore the data. The tree is visualized to show the rules that has been discovered and the predictions are reasonable. To summarize, Duggan supports Edali and Yüsel's perspective and states that the combination of machine learning and simulation models provides a potential for greater insight into the modelling process (Duggan, 2020).

Chen et al. proposed a design method for simulation models where the selected policies were based on neural networks (Chen, Tu and Jeng, 2011). First, the policy maker develops a behavioral pattern. Then an algorithm is used to find the most appropriate model while both the system and the values are changed during that process. This process consists of using a recurrent neural network (RNN) to reformulate the policy design into a machine learning task and then using genetic algorithms (GA) to solve the task. RNN was chosen as it is equivalent to system dynamic models (Chen, Tu and Jeng, 2011). In order to reformulate policy design into a machine learning task, Chen et al. proposed the following;

1. Genetic encoding of the policy design or the RNN that represents it
2. Select a fitness function to evaluate the solution
3. Set up the evolution process to produce the individuals
4. Evaluate the generation and use as result or generate again

Chen et al. tested their approach to several cases, including the *World Dynamics* model developed by Forrester in 1973 (Chen, Tu and Jeng, 2011). The results showed that this procedure gave as good or better results. Therefore, they argue that machine learning, and in specific RNNs, is of good value for policy training. They do not propose to substitute human experts with machine learning, but use it as an assistant to enhance the model development (Chen, Tu and Jeng, 2011).

In 2017, Abdelbari and Shafi studied how computational methods can be used to form system dynamic models directly from observed data (Abdelbari and Shafi, 2017). They propose a neural network based approach to learn causal loop structures. The methodology starts with identification of the most important variables to a system. Then, data is collected that describes the behavior of the system. A population of echo state networks (ESN) are designed, and the best model is selected. An echo state model is a type of RNN. Abdelbari and Shafi suggested that this approach can complement the conceptual model development. The approach was tested on three case studies, where the ESN was able to learn the exact structure of only one. However, with further research and modifications this methodology could possibly be scaled to more complex systems.

In 2022, Weng et al. studied the utilization of machine learning for modelling chaotic systems (Weng *et al.*, 2022). They found that reservoir computing provides an alternative way of modelling these systems, rather than the more traditional dynamic equations. A reservoir computational model consists of three layers; an input layer, a reservoir network and then the output layer (Weng *et al.*, 2022). Weng et al. showed that these models provide the same recurrence properties as dynamical equations, in addition to other benefits. They concluded that reservoir computing can accurately describe chaotic systems.

System dynamics has also been used to cover well-known flaws of machine learning. While there has been exponential growth in the application of machine learning, it has been shown that these techniques might lead to discrimination (Prabhakaran and Martin Jr, 2020). An example is utilization of machine learning within healthcare, where it has been shown that some risk-assessment tools have had racial bias, denying certain groups to special programs and resources. In order to avoid discrimination in such technologies, Prabhakaran and Martin Jr developed a community-based system dynamics method. This method was developed to combine diverse sources of causal series. It was developed by having group sessions with diverse stakeholders and visualizations of proposed causal theories. In addition, computer simulations were used to expose the dynamics of complex problems from a feedback perspective. They believe this approach will improve machine learning fairness and thereby protect human rights (Prabhakaran and Martin Jr, 2020).

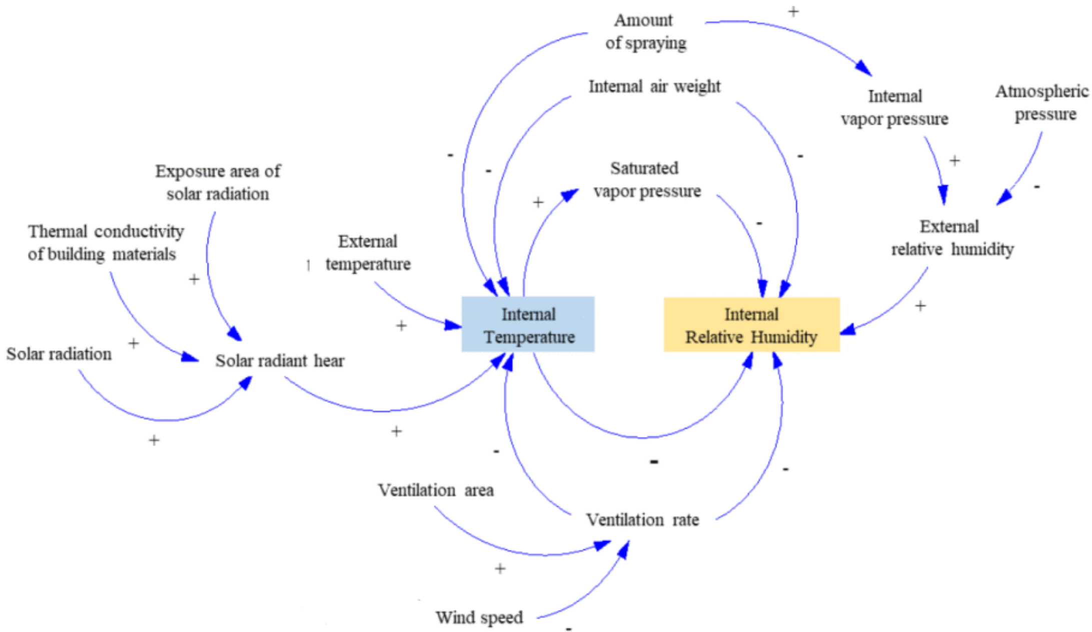


Figure 16: System dynamics model for the greenhouse micro-climate (Chen *et al.*, 2022)

Another application of system dynamics and machine learning is within energy. Greenhouse farming can be considered as a good option in regards to climate change. However, greenhouses do have some challenges regarding their microclimate and rapid weather changes. Chen et al. developed a smart microclimate-control system (SMCS) to regulate the micro-environment of the greenhouse

(Chen *et al.*, 2022). The process began with data collection from Internet of Things (IoT) from the investigated greenhouse. Based on this data, a system dynamics model was built (figure 16). The model simulated the micro-climate before and after spraying. Then a back-propagation neural network model predicted the temperature and humidity one hour ahead. A spray mechanism was designed to act based on these predictions. The SMCS performed as well as the traditional system in regards of environmental control, but additionally comes with other benefits in regards to sustainability and climate.

In 2022, Roozkhosh et al. proposed a new system dynamics and machine learning method to investigate blockchain acceptance rate (BAR) in the home appliances supply chain in Iran (Roozkhosh, Pooya and Agarwal, 2022). Blockchain technology has become popular in supply chain management as it provides security, efficiency, and trustworthiness. The BAR was simulated using system dynamics and will be further developed in the next years based on sensitivity analyses, policy design and validation. Figure 17 shows a subsystem of the model, consisting of personnel skills and how this affects the probability of using blockchain technology. After the simulation, the behavior in the simulated data was further analyzed using machine learning. Multi-Layer Perceptron (MLP) and support vector regression (SVR) was used to predict the BAR behavior. The predictions were evaluated on data from 2020-2022. It was shown that policy design can have beneficial effects for increasing the resilience in the supply chain. They also concluded that the system dynamics model gave better results when combining with MLP than with SVR.

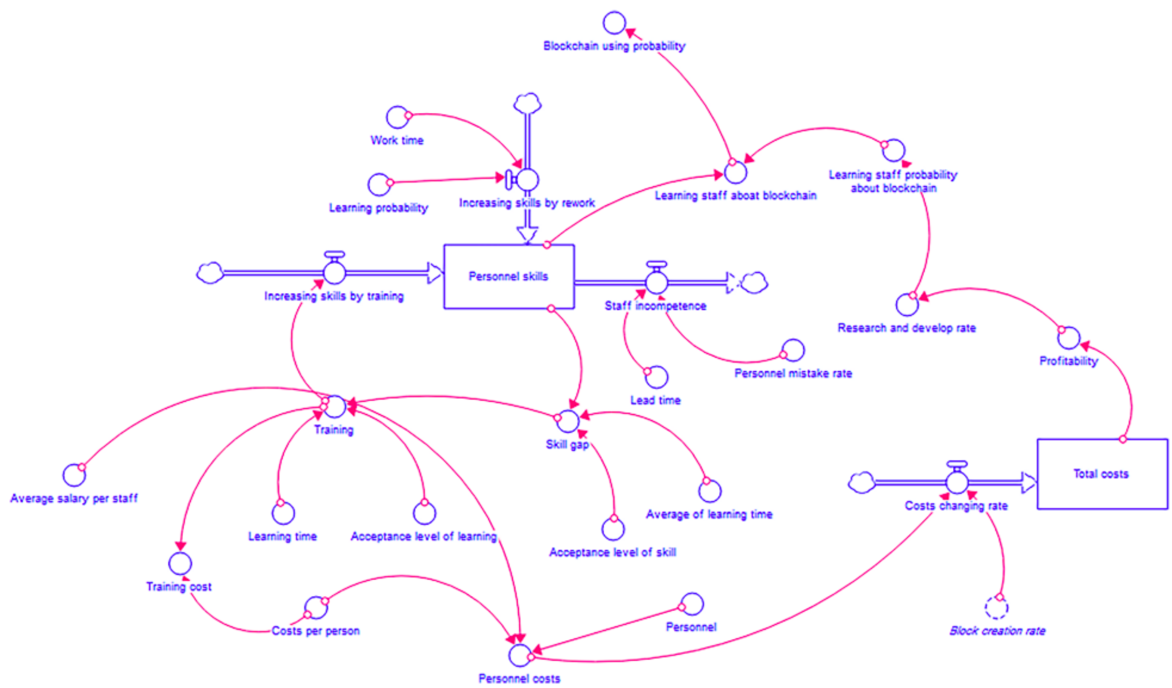


Figure 17: System dynamics model subsystem of personnel skills (Roozkhosh, Pooya and Agarwal, 2022)

Table 6 presents an overview of the articles and their associated machine learning algorithms. In addition to the tree based algorithms decision tree and random forest, it is evident that neural networks and in specific RNNs are commonly used together with system dynamics. The article by Prabhakaran and Martin Jr does not have an algorithm as they write about machine learning in general.

Machine learning algorithms		
Author(s)	Article name	Algorithm(s)
Abdelbari and Shafi	A computational Intelligence-based Method to 'Learn' Causal Loop Diagram-like Structures from Observed Data: Machine-learning of causal-loop diagrams	ESN(RNN)
T.-H. Chen et al.	Develop a Smart Microclimate Control System for Greenhouses through System Dynamics and Machine Learning Techniques	BPNN
Y.-T. Chen et al.	A Machine Learning Approach to Policy Optimization in System Dynamics Models	RNN+GA
Duggan	Exploring the opportunity of using machine learning to support the system dynamics method: Comment on the paper by Edali and Yücel	DT
Edali and Yüsel	Analysis of an individual-based influenza epidemic model using random forest metamodels and adaptive sequential sampling	RF
Prabhakaran and Martin Jr	Participatory Machine Learning Using Community-Based System Dynamics	-
Roozkhosh et al.	Blockchain acceptance rate prediction in the resilient supply chain with hybrid system dynamics and machine learning approach	MLP+SVR
Weng et al.	Modeling chaotic systems: Dynamical equations vs machine learning approach	RNN

Table 6: Overview of algorithms which are combined with system dynamics

3 Methodology

In order to investigate how system dynamics and machine learning can be combined for safety predictions, a simulation model has been developed and machine learning models have been tested. Activities and indicators were chosen based on theory and then modelled through influencing diagrams. The model was implemented and run to generate datasets which were used for machine learning predictions. Hence, this section covers information on the thesis context, the conducted literature reviews, development of the system dynamics model and methodology for machine learning. Additionally, an evaluation of the method is provided in terms of reliability and validity.

3.1 Thesis Context

This subsection provides more information on the context of this thesis, including the associated research method, program and order of operations for combining system dynamics and machine learning. Additionally, a description of how the system dynamics model is developed in collaboration with another master student is presented.

3.1.1 Quantitative Research Method

Generally research is divided into two distinct categories; quantitative and qualitative methods. Some argue that quantitative methods and qualitative methods are not polar opposites but different ends of a spectrum (Newman, Benz and Ridenour, 1998). Quantitative research involves examining the relationship between variables to test objective theories. Measurable variables in quantitative research allow for the analysis of numerical data using statistical procedures (Creswell, 2009). In quantitative approaches, measurements are made through gathering data (Fellows and Liu, 2003). Qualitative research uses methods comprehending and examining the significance that individuals or groups attribute to a social or human situation (Creswell, 2009). These methods may make it easier to appreciate and comprehend fundamental causes, ideas, and behaviors (Fellows and Liu, 2003). If a qualitative research methodology is chosen, then possible methods could include conducting a case study, a personal interview, or making observations (Lowhorn, 2007). Qualitative research strives to build a theory that explains observed behavior, whereas quantitative research seeks to validate a theory through experiments and numerical analysis of outcomes. Therefore, it is believed that qualitative research is more inductive while quantitative research is more deductive (Lowhorn, 2007).

The choice between quantitative or qualitative methods will be dependent on the outcome of the research. The current situation as well as the research questions should be taken into consideration to make an informed decision (Lowhorn, 2007). In the context of this thesis, the need for numerical data was identified for both the system dynamics model and for machine learning. However a significant amount of relevant project data was qualitative. Creating a simulation of a construction project meant qualitative data needed to be made quantitative to be used in the model. One example of this was measuring the quality of a SHA plan with a number from one to ten. Apart from this adaptation of qualitative data, the remainder of the methods used in this thesis predominantly followed a quantitative approach. This involved the collection of a large volume of numerical data from the system dynamics model, which was subsequently utilized for machine learning purposes. Related to General Data Protection Regulation (GDPR), synthetic data does not pose a GDPR concern since it does not contain any personal data. It was therefore not necessary to notify the Norwegian centre for research data.

3.1.2 DiSCo Project

This master thesis is written as a contribution to the Sustainable value creation by digital predictions of safety performance in the construction industry (DiSCo) project. The research project is a collaboration between NTNU, Skanska, Sporveien, Norconsult and Safetec. The project is funded by the Research Council of Norway. It is led by the Department of Industrial Economics and Technology Management at NTNU.

The purpose behind the DiSCo project is to develop knowledge and methods for utilization of AI in the early phases of construction projects in order to predict safety performance during production (NTNU, n.d.). This can be a contribution to an improved decision-making support to reduce the number of accidents within the industry. The DiSCo project will give a better understanding on how different machine learning techniques can be used on available data to gain innovative and proactive safety management methods. The research involves which early factors influence safety during production, how machine learning can be used to give early warnings, and demonstrates how this can lead to fewer accidents.

3.1.3 Combining System Dynamics and Machine Learning

This research aimed to combine system dynamics and machine learning to investigate early warnings of projects with a high accident risk. The reasons behind combining these two technologies were a lack of relevant project data and the opportunity for exploration. The development of the system dynamics model allowed for creation of synthetic data which could be used for machine learning predictions. The research would then serve as a proof of concept on how safety factors influence the number of accidents and how machine learning could be utilized for accident prediction.

First, the system dynamics model was developed. This model allowed for creating data on all relevant phases of a construction project, including both planning- and the construction phase itself. The data was generated into numerous indicators, representing different processes and deliveries through the project. These indicators were based on theory and previous work, and can be difficult to gather real data on. The system dynamics model also allowed for Monte Carlo simulations with uncertainties such that different variations of outcomes were generated. The simulations could be run as many times as wanted, such that one would be able to choose the length of the dataset.

After the dataset had been generated by the simulation model, it was utilized for machine learning predictions. The features from the construction phase, except from the target feature being number of accidents, were removed from the dataset. This was in order to examine whether information from only the planning phase would be enough for accurate predictions of safety performance. Then machine learning algorithms were used to predict safety performance of the projects in the dataset. The process is visualized in figure 18.

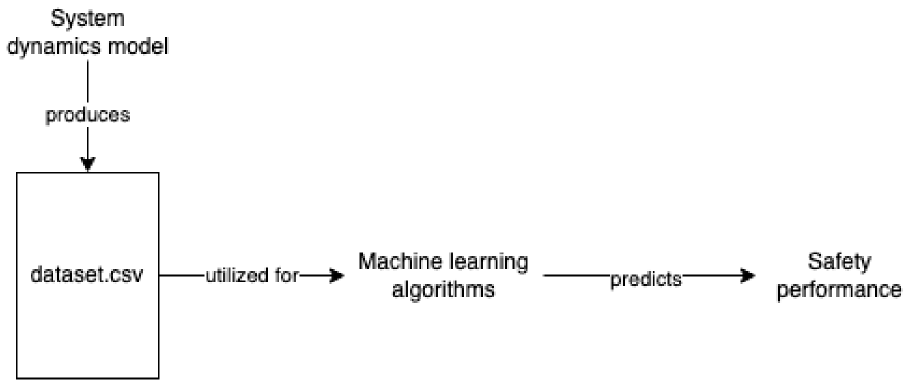


Figure 18: Utilization of system dynamics and machine learning

3.1.4 Planning- and Construction Phase Models

The development of the system dynamics model was done in collaboration with master student Josefine Stiff Aamlid. The model was separated into two parts; the planning phase and the construction phase. This thesis focused on the planning phase. Information regarding the construction phase can be found in *System Dynamics Modelling and Machine Learning to Improve Safety in Construction Projects* (Aamlid, 2023).

There was a tight communication through development of the system dynamics model. Weekly meetings were held in addition to smaller stand-up meetings if needed. The planning and construction phases were developed in parallel. GitHub was used for collaboration and version control. To be able to connect the models, it was discussed at an early phase which indicators from the planning phase should be used as input for the construction phase. These are presented in section 3.3.2. Arbitrary values were used as input in the meantime.

When both sub-models had a first draft developed, they were put together. In practice this was done by generating a file with all indicators from the planning phase. The construction phase script then read the file and selected the indicators which were used as input for some of its own functions. Then the simulation was run for the construction phase as well, and simulated accidents were produced. A new file, the *dataset*, was generated from this execution, containing all data from both the planning and construction. This was done with both serious and fatal accident rates, producing two separate datasets. The process is shown in figure 19.

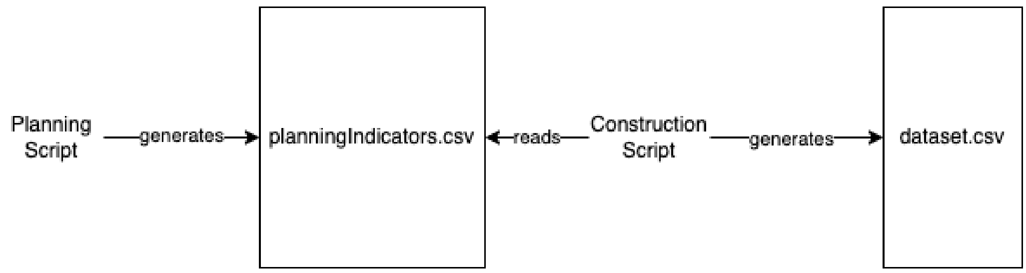


Figure 19: Development of the system dynamics model

Despite the simulation model being developed in collaboration, different datasets were used for machine learning predictions in the two theses. The same algorithms and evaluation metrics were selected, but implemented with different parameters. This yielded different machine learning results between the theses.

3.2 Literature Review

Four literature reviews were conducted for the theoretical background. The search engine Oria was used for the reviews. Oria provides access to the different databases from Norwegian university libraries. It allows for filtering on the articles that have been peer reviewed, which guarantees that the papers are authored and reviewed by professionals. Information regarding each review is provided below.

3.2.1 Safety Factors in Construction Projects

This literature review was done to uncover previous research on which factors influence safety performance in construction projects. The findings were presented in section 2.2.2. Table 7 shows the search queries that were used.

Search Query	No. of hits
safety AND factors AND performance AND "construction projects"	477
"safety factors" AND performance AND "construction projects"	25
safety AND factors AND performance AND "construction projects" AND (affecting OR influencing)	134

Table 7: Search queries for safety factors in construction projects

As the first query gave very many hits, it was decided to narrow it down. Therefore the second query attempted to underline that safety factors was the point of research. However, some relevant articles seemed to have disappeared by this change. A third query included the words affecting or influencing. This gave 134 hits, which was reasonable. The following selection criteria were applied (table 8);

Id	Criteria
IC1	Full text available
IC2	Of type “article”
IC3	The document is in English
IC5	Peer reviewed
EC1	Published before 2018

Table 8: Selection criteria for safety factors

This resulted in 68 articles. They were reviewed on title and nine articles were selected. The selected articles were the ones that seemed to have their main focus on factors.

3.2.2 System Dynamics and Safety Performance

A literature review was conducted to see how system dynamics had been used previously to simulate safety performance in construction projects. The findings were presented in section 2.3.3. The search queries are presented below (table 9).

Search Query	No. of hits
system dynamics AND safety performance	22 218
“system dynamics” AND safety performance	972
“system dynamics” AND “safety performance”	57
“system dynamics” AND “safety performance” AND (project OR construction)	23

Table 9: Search queries for system dynamics and safety performance

In the beginning it was clear that not using the search terms in brackets returned too many results. Finally, the search resulted in 23 hits where 21 results remained after the inclusion criteria were applied (table 10).

Id	Criteria
IC1	Full text available
IC2	Of type “article”
IC3	The document is in English
IC5	Peer reviewed

Table 10: Inclusion criteria for system dynamics and safety performance

3.2.3 Machine Learning for Safety Predictions

This literature review was conducted in connection to our specialization report (Borkenhagen and Olsen, 2022). For the master thesis, several of the articles were reviewed again to elaborate on their objectives. The literature review done for the specialization project is explained below.

Search Query	No. of hits
(“machine learning” OR AI) AND accident	5 046
(“machine learning” OR AI) AND accident AND construction	496
(“machine learning” OR AI) AND “accident prediction” AND construction	29

Table 11: Search queries for machine learning for safety predictions (Borkenhagen and Olsen, 2022)

Table 11 shows the search queries for safety and machine learning. This search was also conducted through Oria. Based on the last query (table 12), 29 articles were reviewed on title and abstract. This resulted in 22 relevant articles.

3.2.4 System Dynamics and Machine Learning

A literature review was conducted for the combination of system dynamics and machine learning. The search queries and number of hits are shown in table 12.

Search Query	No. of hits
“system dynamics” AND ”machine learning”	1236
title: “system dynamics” AND ”machine learning”	44

Table 12: Search queries for system dynamics and machine learning

The first search query resulted in a large number of hits and it became evident that those covered a very wide spectrum. To narrow it down and make sure the results were mainly focused on system dynamics and machine learning, another search was conducted where the title had to include those keywords. This resulted in a more manageable number of hits, while they seemed to be more relevant. It was detected how Oria’s advanced search on title included all front page titles including the article title. Next, several selection criteria were applied. These were the same as presented in table 10.

After the application of the inclusion criteria the remaining hits were 25 articles. These were reviewed based on title and abstract. The number of articles were reduced to 22 due to duplicates. These 22 articles were categorized into seven groups; general, health, energy, operations management, food production, transport and climate. The articles were further reviewed categorically, while 15 of them were considered outside the thesis scope. In many of the cases, this was due to the “system dynamics” word combination being used to describe other systems and not system dynamic models as they are defined in this thesis. This led to seven relevant articles, with one of them commenting another relevant article which was also added. As a result, 8 articles were selected.

3.3 System Dynamics Model

As presented in section 3.1.4, this thesis evolves around the planning phase of construction projects. Hence, this section provides information on how the system dynamics model for the planning phase was created. This includes selection of stages and activities, selection and dependencies of indicators and the technical implementation of the model.

For simplification, it was decided to narrow down to a certain type of project; "green-field" construction of buildings. Design-build contracts were chosen as contractual arrangement. Additionally, contract type was set to fixed price contracts.

3.3.1 Planning Activities

When simulating the planning phase of a project the different activities within each stage needed to be defined. The four main stages seen in figure 20 are based on Neste Steg and are identical to the ones presented in the theoretical background (figure 3). Further, the activities within each stage were defined based on the deliveries from the theoretical findings (section 2.1.3). Figure 20 shows how these activities are connected.

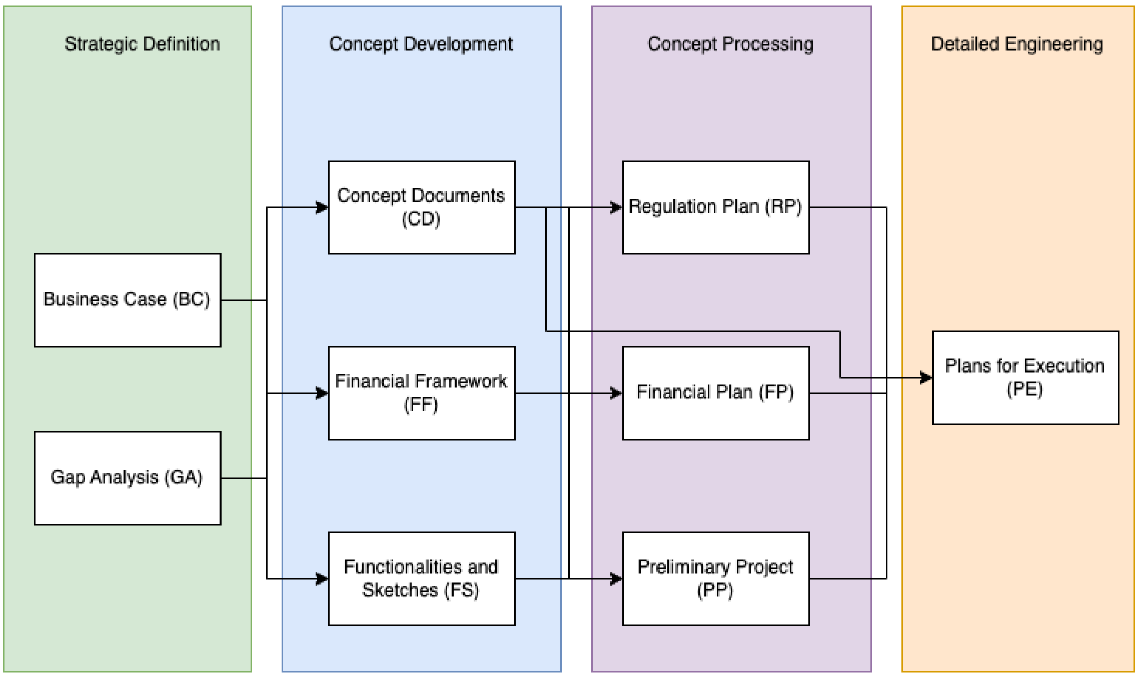


Figure 20: Activities in the planning phase

The first stage is the *Strategic Definition* which includes the *Business Case* and the *Gap Analysis*. The business case activity is the part of the project where the goals and objectives are defined. A business plan is created to make sure the construction project is aligned with the goals and objectives. This includes a thorough examination of the project's risks and challenges. The gap analysis activity is focused on identifying the current state and the desired end state. The differences between these states is then evaluated to identify the gap. These activities are executed simultaneously and both must be completed before the project may proceed to the following part of the planning phase.

The next three activities are *Concept Documents*, *Financial Framework* and *Functionalities and Sketches*. This part of the planning phase is based around developing the concept further and is called *Concept Development*. The concept documents activity is the foundation for creating more precise planning and design for the construction project. Additionally, it creates a shared understanding of the goals and methods to be used, which can lower risk and avoidable costs later on in the project. The guidelines and standards for handling the project's finances are contained in the financial framework activity. It includes the budget for the entire project. The functionalities and sketches activity includes early design concepts and drawings that examine the fundamental form and layout of a building as well as a description of the functions required to satisfy the demands and requirements of the structure's users. This will serve as the foundation for more detailed design work later on in the project.

The next stage of the planning phase is *Concept Processing* and includes three activities. These activities are *Regulation Plan*, *Financial Plan* and *Preliminary Project*. The regulation plan activity revolves around how the project manages the different regulations from the government. To ensure that all regulations and standards are accurately understood and adhered to, coordination with authorities is crucial. The financial plan activity is a more detailed version of the financial framework activity. This activity includes a more comprehensive budget where detailed cost estimates are established. A portion of this more detailed budget will be set aside for safety-related initiatives. The strategy for the project's financing and contract structure will also be included. The preliminary project activity is further development of the project and contains more details about the chosen concept. Calculations go through quality assurance and drawings are finalized. Solutions are controlled to ensure that the project can be realized.

The final stage of the planning phase is *Detailed Engineering*. This stage contains only one activity and that is *Plans for Execution*. The plans for execution activity contains thorough technical drawings and descriptions of how the project will be carried out. This stage of a construction project is crucial because precision and quality are necessary to ensure that it can be completed effectively and securely. Technical specifications preparation, equipment and material selection, and descriptions of work procedures are all included in this activity.

3.3.2 Indicators

Within each activity different indicators were chosen. Indicators were selected based upon what indicators people have used previously (section 2.3.3) and from theoretical research (section 2.1 and 2.2). A list of the indicators chosen can be seen in table 13.

Indicator	Domain	Unit
ProjectSchedule	Integer	Days
ByggherreCompetency	[1,10]	
EntreprenørCompetency	[1,10]	
BusinessCase.ProjectComplexity	[1,10]	
BusinessCase.ProjectSize	[1,10]	
BusinessCase.ProjectExpectedStartDate	Datetime	
BusinessCase.ProjectExpectedEndDate	Datetime	
PlanningDuration	Integer	Days
ConstructionDuration	Integer	Days
BusinessCase.ExpectedDuration	Integer	Days
BusinessCase.ActualDuration	Integer	Days
BusinessCase.RiskAssessment	[1,10]	
BusinessCase.Quality	[1,10]	
GapAnalysis.ExpectedDuration	Integer	Days
GapAnalysis.ActualDuration	Integer	Days
GapAnalysis.StakeholderReq	[1,10]	
GapAnalysis.Quality	[1,10]	
ConceptDocuments.ExpectedDuration	Integer	Days
ConceptDocuments.ActualDuration	Integer	Days
ConceptDocuments.ProjectStatus	Integer	Days
ConceptDocuments.Quality	[1,10]	
FinancialFramework.ExpectedDuration	Integer	Days
FinancialFramework.ActualDuration	Integer	Days
FinancialFramework.ProjectStatus	Integer	Days
FinancialFramework.Budget	Integer	MNOK
FinancialFramework.Quality	[1,10]	
FunctionalitiesSketches.ExpectedDuration	Integer	Days
FunctionalitiesSketches.ActualDuration	Integer	Days
FunctionalitiesSketches.ProjectStatus	Integer	Days
FunctionalitiesSketches.Quality	[1,10]	
RegulationPlan.ExpectedDuration	Integer	Days

RegulationPlan.ActualDuration	Integer	Days
RegulationPlan.ProjectStatus	Integer	Days
RegulationPlan.CoordinationAuthorities	[1,10]	
FinancialPlan.ExpectedDuration	Integer	Days
FinancialPlan.ActualDuration	Integer	Days
FinancialPlan.ProjectStatus	Integer	Days
FinancialPlan.SafetyBudget	[1,10]	
FinancialPlan.Quality	[1,10]	
PreliminaryProject.ExpectedDuration	Integer	Days
PreliminaryProject.ActualDuration	Integer	Days
PreliminaryProject.ProjectStatus	Integer	Days
PreliminaryProject.SHAPlanQuality	[1,10]	
PreliminaryProject.Quality	[1,10]	
PlansExecution.ExpectedDuration	Integer	Days
PlansExecution.ActualDuration	Integer	Days
PlansExecution.ProjectStatus	Integer	Days
PlansExecution.LevelOfDetail	[1,10]	
PlansExecution.SchedulePressure	[1,10]	
PlansExecution.SafetyTraining	[1,10]	
PlansExecution.SafetySupervision	[1,10]	
PlansExecution.EquipmentLiability	[1,10]	
PlansExecution.Quality	[1,10]	

Table 13: Selected indicators

Within each activity different indicators were identified. An influencing diagram was created for each activity. The diagrams illustrate the interconnections between indicators, represented by arrows. A plus symbol indicates a positive relationship between the indicators and a minus symbol indicates a negative relationship. The diagrams are variants of causal loop diagrams as presented in section 2.3.1, but also contains the red star symbol, which indicates a more complex relationship between indicators.

In figure 21 the influencing diagram for the business case activity is shown. It can bee seen that the *ExpectedProjectStartDate* affects *ExpectedProjectEndDate* and *ProjectSchedule* is also affected by the size and complexity of the project. In this model, the size is a measure of cost, while complexity describes the tecnical complexity of the project. It is assumed that the *PlanningDuration* is roughly half of the duration of the project in total. The indicator *BusinessCase.ExpectedDuration* was set to be one sixth of the total planning duration. The *BusinessCase.ActualDuration* is the measure of if an activity finished ahead time, on time or behind schedule. This was based on the expected duration and the competency level of the owner.

The indicator *BusinessCase.RiskAssessment* describes the quality of the project's risk assessment. This quality is affected by the complexity of the project where it is assumed that higher complexity makes it more difficult to create a high quality risk assessment. For the risk assessment the indicator *ByggherreCompency* is also important. An owner with low competency is assumed to be more likely to create low quality risk assessments. Lastly the quality of the business case as a whole is represented in the indicator *BusinessCase.Quality*. This indicator is affected by the actual duration and the expected duration. It is also based on the quality of the risk assessment and the competency of the owner.

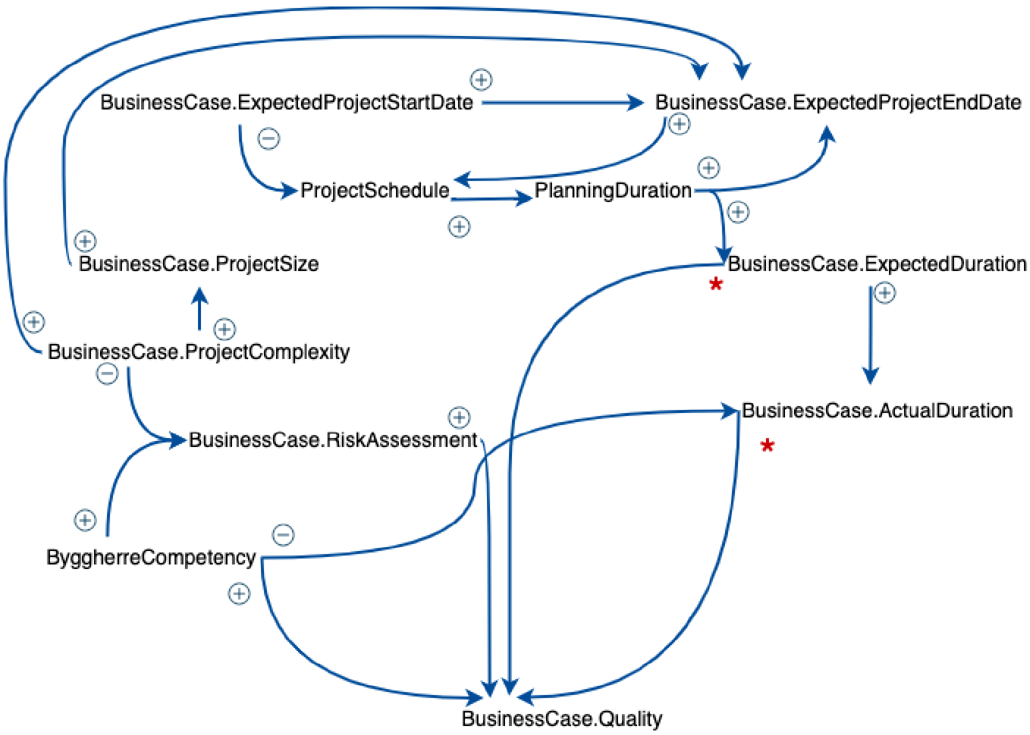


Figure 21: Influencing diagram for the Business Case activity

Figure 22 below shows the gap analysis. The indicator *GapAnalysis.ExpectedDuration* is assumed to be one eighth of the total planning duration given by the indicator *PlanningDuration*. *GapAnalysis.StakeholderReq* is on a scale from one to ten, where a low score indicates a small number of stakeholders with few requirements, and a high score indicates a large number of stakeholders with numerous requirements. A high *GapAnalysis.StakeholderReq* is assumed to give a higher *GapAnalysis.ActualDuration*. The relationship between *GapAnalysis.ExpectedDuration* and *GapAnalysis.ActualDuration* is assumed to affect the quality of the gap analysis activity. The *GapAnalysis.Quality* is therefore based on the difference between the expected and actual duration together with the competency of the owner.

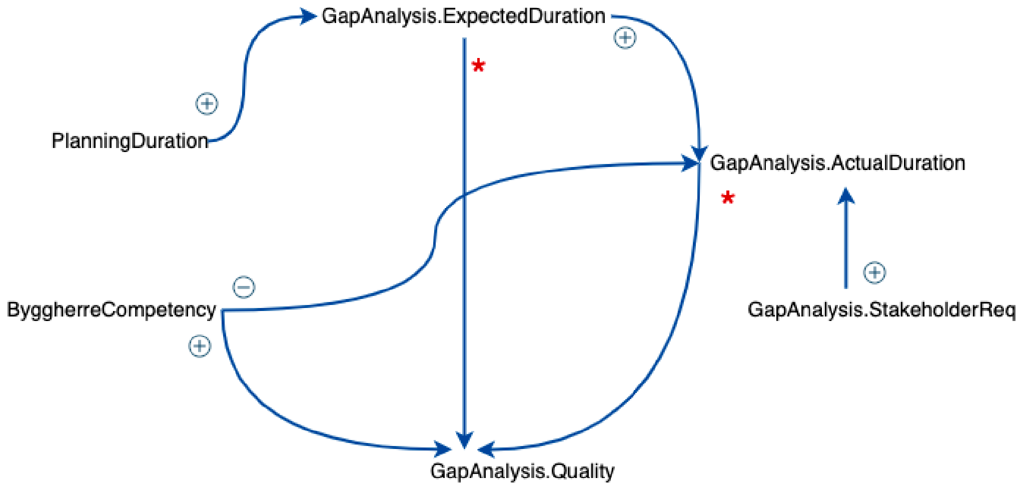


Figure 22: Influencing diagram for the Gap Analysis activity

In figure 23 the influencing diagram for the concept documents activity is shown. It has many similar indicators with the gap analysis activity with the *ConceptDocuments.ExpectedDuration* being set as one third of the total planning duration. The *ConceptDocuments.ActualDuration* is affected by the expected duration and the competency of the owner. The difference between expected and actual duration is given by the indicator *ConceptDocuments.ProjectStatus* which is the amount of days an activity finished either behind or ahead of schedule. Similarly to the gap analysis activity the indicator *ConceptDocuments.Quality* is based on the difference between the expected and actual duration and the competency of the owner. However, the quality is also affected by the quality of the prior two activities. It is assumed that if the indicators *BusinessCase.Quality* and *GapAnalysis.Quality* have high values then the *ConceptDocuments.Quality* indicator will also have a higher value.

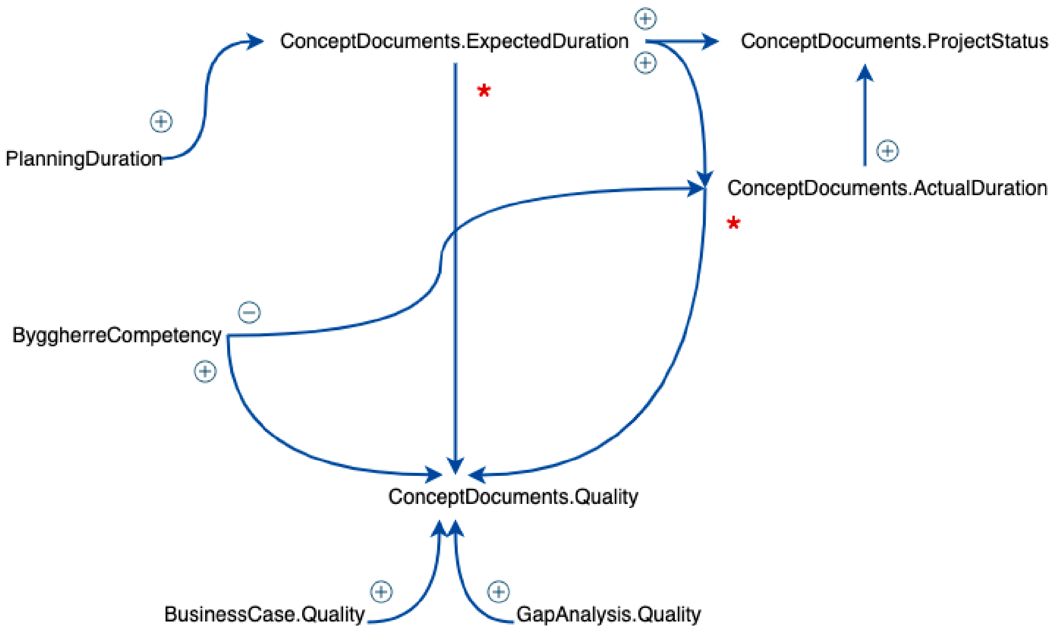


Figure 23: Influencing diagram for the Concept Documents activity

The financial framework diagram is shown in figure 24. The diagram has many similarities with the concept documents diagram. The indicator *FinancialFramework.ExpectedDuration* is assumed to be one fourth of the total planning duration. The indicators *FinancialFramework.ActualDuration*, *FinancialFramework.ProjectStatus*,*ByggherreCompetency*, *FinancialFramework.Quality*, *BusinessCase.Quality* and *GapAnalysis.Quality* affect each other similarly as in the the concept documents indicator diagram. On the left side of the model it can bee seen that the indicator *FinancialFramework.Budget* is affected by *BusinessCase.ProjectComplexity* and *ProjectSchedule*. It is assumed that a higher project complexity and a longer project schedule all lead to a larger budget. Additionally, the project schedule originates from project size, which was an initial measure in terms of cost. The budget is measured in millions of Norwegian kroner (MNOK) and can vary between 96 MNOK and 1440 MNOK.

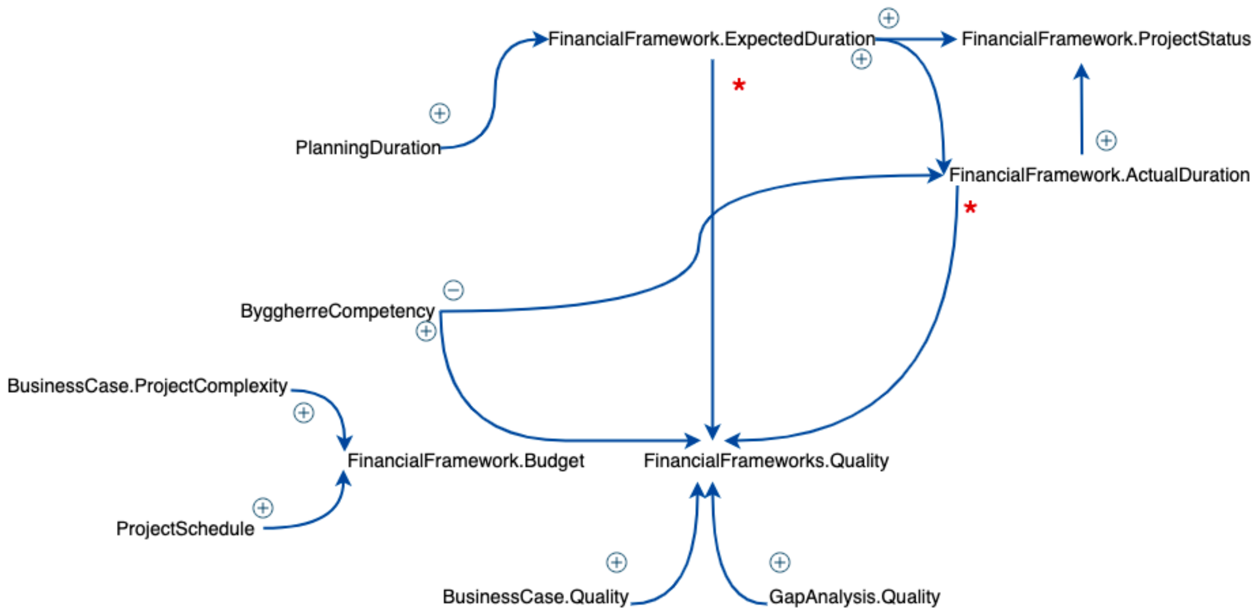


Figure 24: Influencing diagram for the Financial Framework activity

The influencing diagram for the functionalities and sketches activity is shown in figure 25. The *FunctionalitiesSketches.ExpectedDuration* indicator is assumed to be one ninth of the total planning duration. The *FunctionalitiesSketches.ActualDuration* is set based on the expected duration and the competency of the owner. The relationship between the expected and actual duration is what determines *FunctionalitiesSketches.ProjectStatus*. Both *GapAnalysis.Quality* and *BusinessCase.Quality* are important in determining the *FunctionalitiesSketches.Quality*. This quality is also affected by the expected duration, actual duration and the competency of the owner.

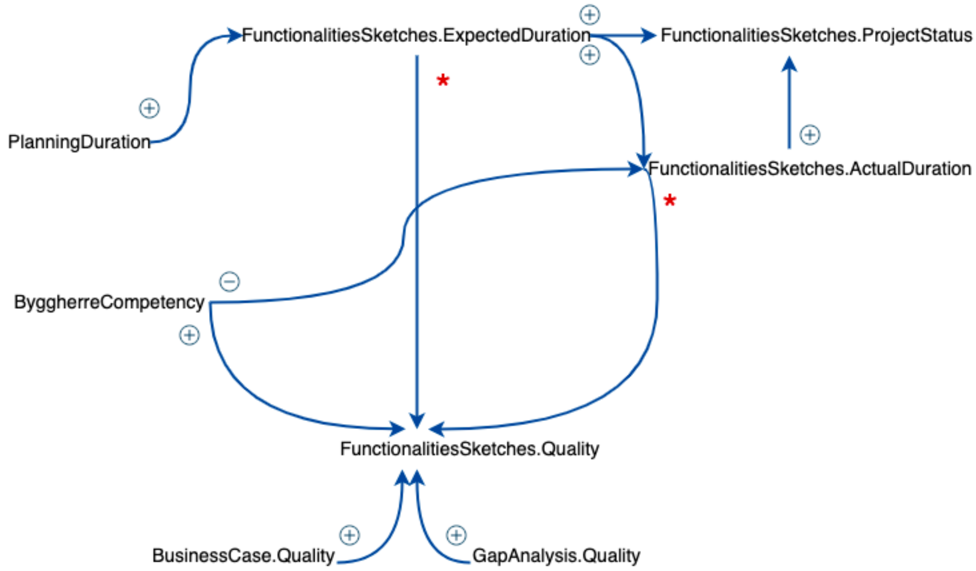


Figure 25: Influencing diagram for the Functionalities and Sketches activity

Figure 26 shows the regulation plan activity. The indicator *RegulationPlan.ExpectedDuration* is set to be one eighth of the planning duration. The indicator *RegulationPlan.CoordinationAuthorities* is a measure on how well the contact between the project management and local authorities has been. A high number for this indicator suggests that all building permits have been granted without further delay and that there have been no misunderstandings regarding laws and regulations with the relevant authorities. It is assumed that a high score on *RegulationPlan.CoordinationAuthorities* means a lower chance of a project being delayed. The *RegulationPlan.ExpectedDuration* indicator is therefore affected by the coordination with authorities and the competency of the owner. The indicator *RegulationPlan.ProjectStatus* keeps track of the project's progress and whether it is currently behind or ahead of schedule. This indicator therefore needs to consider the indicators *ConceptDocuments.ProjectStatus*, *FinancialFramework.ProjectStatus* and *FunctionalitiesSketches.ProjectStatus* as well as how this current activity finished according to schedule. It is assumed that one activity can not begin before the previous activities have finished and therefore the project status with the most delay will be the one used for determining project status.

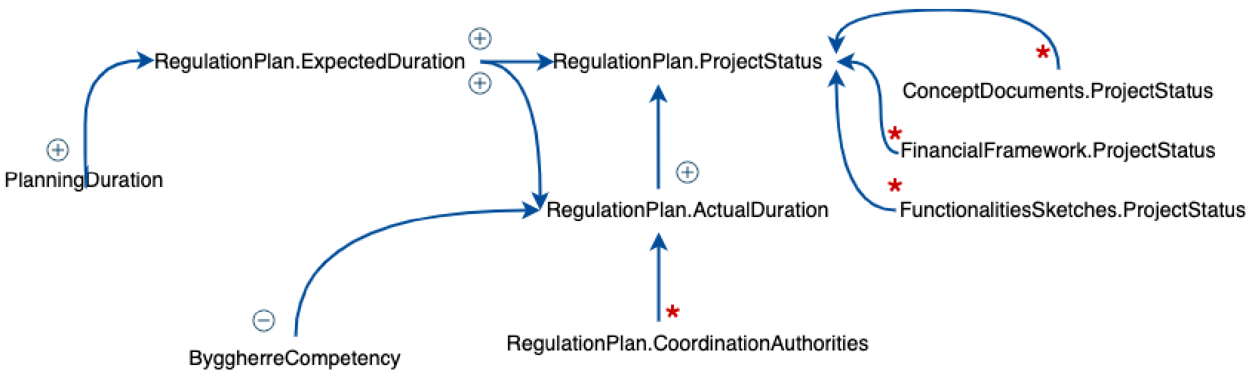


Figure 26: Influencing diagram for the Regulation Plan activity

Figure 27 shows the financial plan activity. The *FinancialPlan.SafetyBudget* is an indicator of how much money is set aside for safety-related costs. A high number means a relatively large safety budget. *FinancialPlan.ExpectedDuration* is set to be one eighth of the planning duration. The *FinancialPlan.EffectiveRiskManagement* is based on *ByggherreCompetency* and *ConceptDocuments.Quality*. Effective risk management is a measure of how well risks are analysed and appropriate strategies are made to proactively reduce these risks. The *FinancialPlan.ProjectStatus* is just like the regulation plan activity based on the project status of previous activities as well as the actual duration of the financial plan activity. The *FinancialPlan.Quality* indicator is based on the *FinancialPlan.EffectiveRiskManagement* indicator and the *FinancialFramework.Quality* indicator.

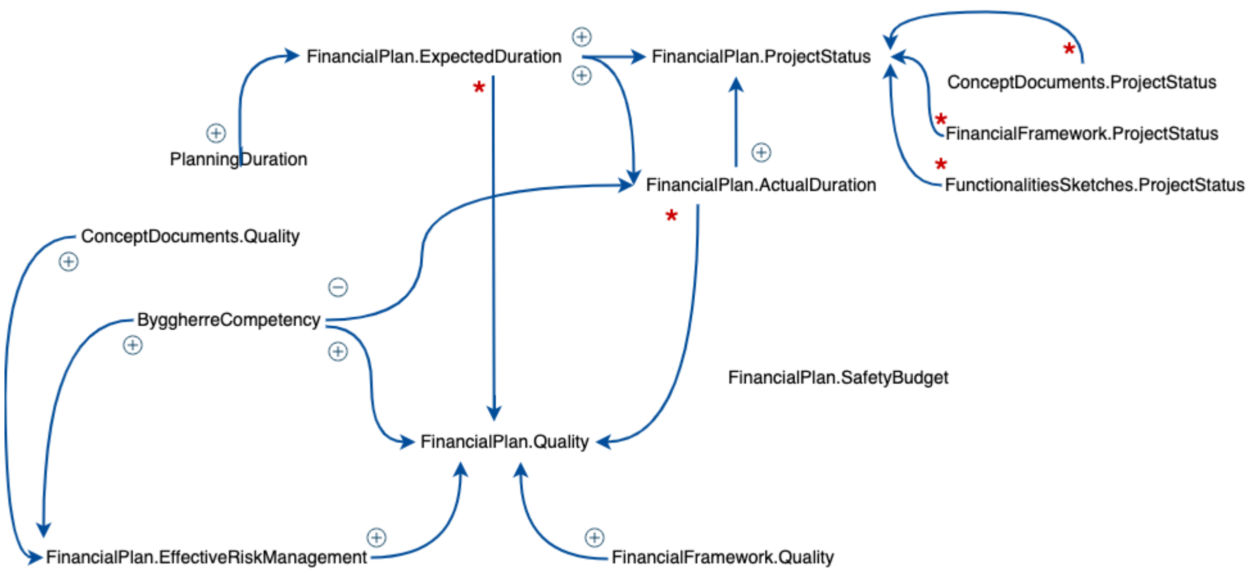


Figure 27: Influencing diagram for the Financial Plan activity

The preliminary project activity can be seen in figure 28. The *PreliminaryProject.ExpectedDuration* is set to be one sixth of the planning duration. The *PreliminaryProject.ProjectStatus* indicator is set based on the project status of previous activities and how the preliminary project finished according to schedule. This is done by adding the difference between the expected duration and the actual duration and adding the largest delay from previous activities. This makes the project status accumulative through the different parts of the planning phase. The indicator *PreliminaryProject.SHAPlanQuality* is an evaluation of the SHA plan's level of quality from one to ten. A detailed and comprehensive SHA plan will score highly on this indicator. It is assumed that the competency of the owner and the quality of the concept documents activity will have an impact on the quality of the SHA plan. The *PreliminaryProject.Quality* is affected by the quality of the functionalities and sketches activity, the competency of the owner and the relationship between the expected and actual duration.

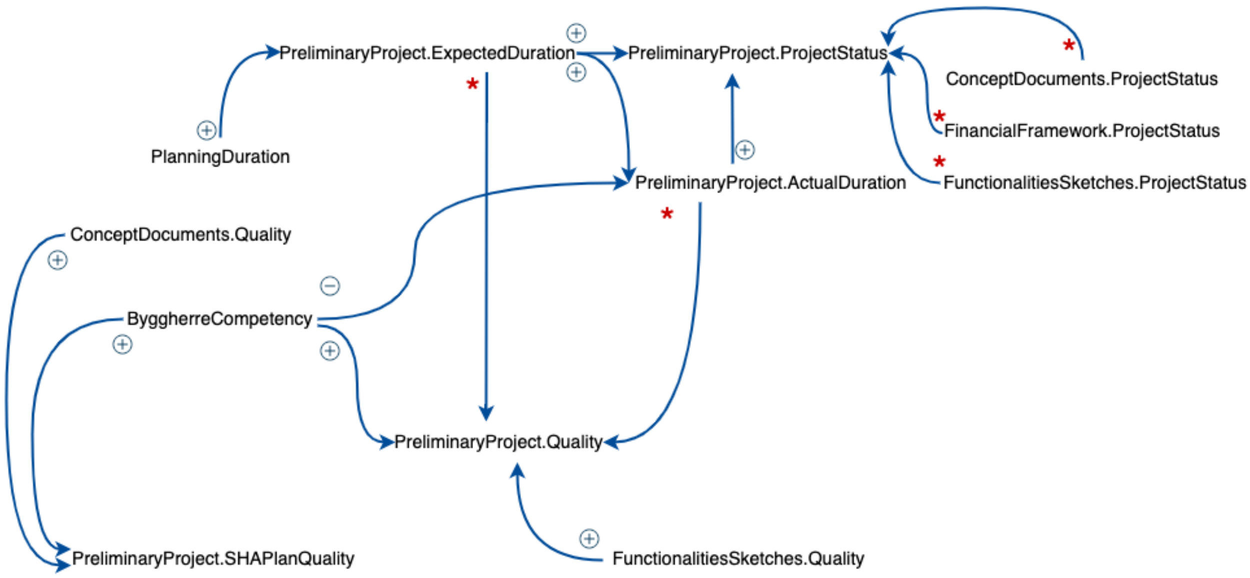


Figure 28: Influencing diagram for the Preliminary Project activity

Figure 29 shows the influencing diagram for the plans for execution activity. The *PlansExecution.EquipmentLiability* indicator is a measure of how good the quality on the project's equipment is. Faulty or old equipment will result in a low score. The indicator *PlansExecution.SchedulePressure* is a measure of how much pressure the project is experiencing due to its schedule. This indicator is based on the project schedule, the planning duration, the construction duration as well as the budget and the *PlansExecution.ProjectStatus*. The *PlansExecution.ExpectedDuration* is half of the planning phase. This makes this activity the longest activity in the planning phase. The *PlansExecution.ProjectStatus* is calculated based on *FinancialPlan.ProjectStatus*, *RegulationPlan.ProjectStatus* and *PreliminaryProject.ProjectStatus*.

The two indicators *PlansExecution.SafetyTraining* and *PlansExecution.SafetySupervision* are both assumed to be based on the *FinincialPlan.SafetyBudget* and the *EntrepenørCompetency*. The *EntrepenørCompetency* indicator is a measure of the contractor's competency on a scale from one to ten. The competency of the contractor is assumed to affect the *PlansExecution.LevelOfDetail* indicator. It is assumed that a contractor with a high competency would more often produce more detailed plans for the execution of the project. It is also assumed that a higher level of detail could cause the project to use more time and therefore could cause an increase in *PlansExecution.ActualDuration*. The *PlansExecution.Quality* indictor is affected by the *EntrepenørCompetency*, the *FinincialPlan.Quality* and the *PreliminaryProject.Quality*. It is assumed that high scores on these indicators lead to a high *PlansExecution.Quality* score. The relationship between *PlansExecution.ExpectedDuration* and *PlansExecution.ActualDuration* is also assumed to affect the quality.

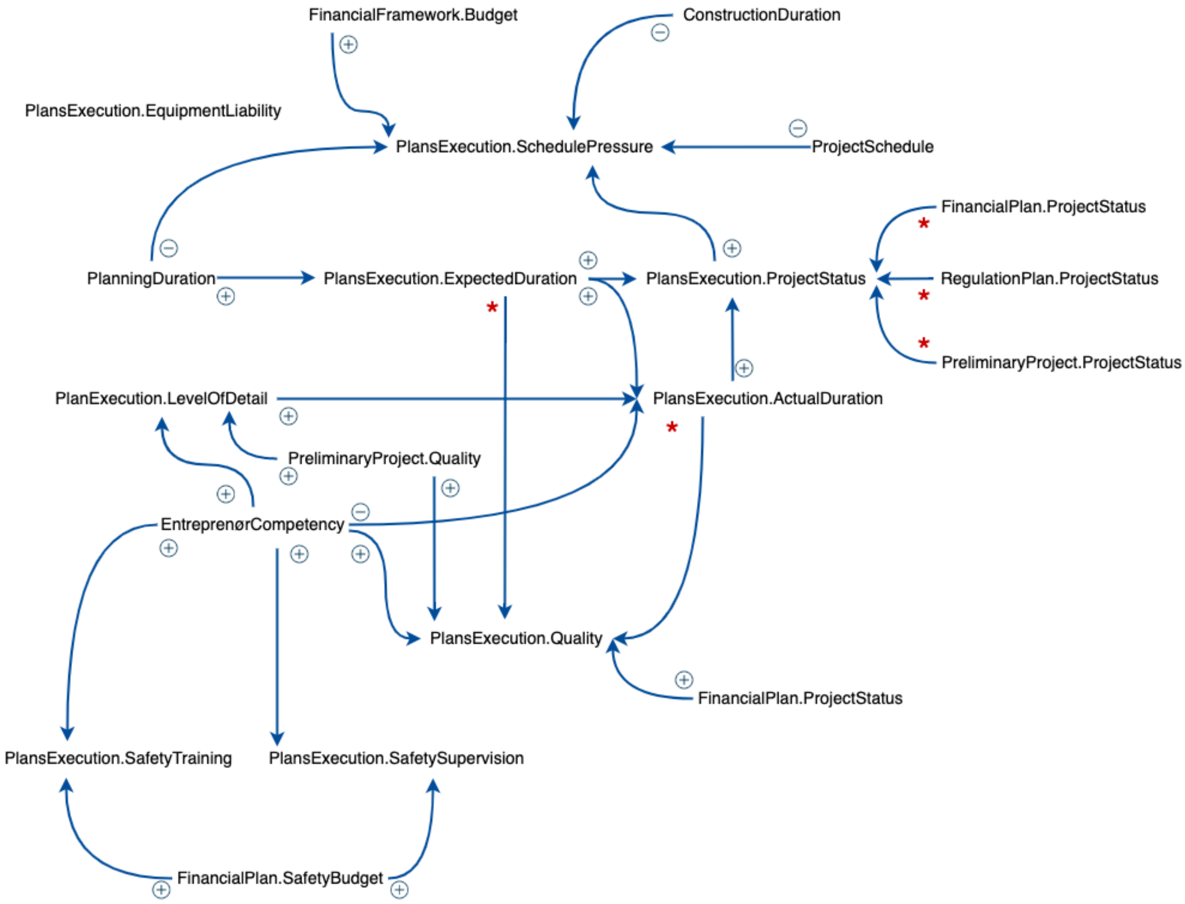


Figure 29: Influencing diagram for the Plans for Execution activity

Eight of the presented indicators were used as input for the construction phase indicators and therefore connecting the two models. These are shown in table 14. They were selected according to the research done for the construction phase (Aamlid, 2023).

Indicator	Domain	Unit
ProjectSchedule	Integer	Days
EntreprenørCompetency	[1,10]	
ConstructionDuration	Integer	Days
PlansExecution.SchedulePressure	[1,10]	
PlansExecution.SafetyTraining	[1,10]	
PlansExecution.SafetySupervision	[1,10]	
PlansExecution.EquipmentLiability	[1,10]	
PlansExecution.Quality	[1,10]	

Table 14: Planning indicators as input for construction indicators

3.3.3 Implementation

The simulation model was developed using the programming language Python. As the model was to consist of projects with associated activities and safety indicators, it was decided to program object-oriented. Classes were defined for projects, tasks and indicators. The associated hierarchy is shown in figure 30. Each box represents a class with associated attributes and operations. As described in the figure, each project can have multiple tasks and indicators, whereas each task and indicator can only belong to one project. In order to simulate multiple projects, the MonteCarloSimulation function is called which runs the project simulation multiple times with the indicators being reset between each run. Due to the uncertainty in the real world and the model, several indicators are set with more random values. When the program is executed, each project iterates through it's tasks dictionary. For every task, an action is performed. This is a function that adjusts certain indicators. When the project has run through the entire planning phase, the indicators are being reset for the next iteration. This process is repeated a predefined number of times. For the final simulations, it was decided to generate 10 000 projects per dataset.

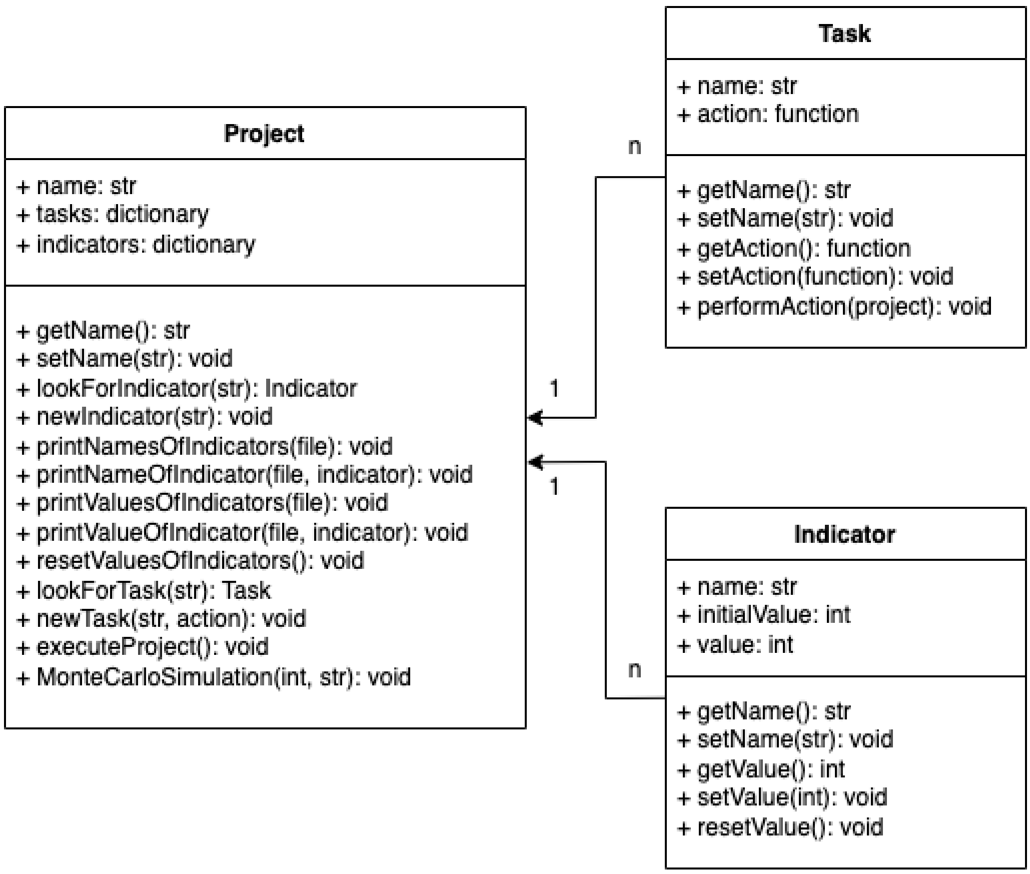


Figure 30: Class diagram for the system dynamics model

The structure described above allows for extension of the model as one can easily develop different kinds of projects with varying tasks and indicators. As previously described in the methodology, only one type of project was simulated in this master's thesis. This was in order to simplify the model. As there is implemented uncertainty in addition to connections between the indicators, the project will be simulated with varying values for each run. With this approach, the project structure remains even though there are varying values being executed.

After the structure was implemented, the selected type of project with its associated tasks and indicators was instantiated. For the planning phase, this was based on the activities and indicators represented in section 3.3.1 and 3.3.2. Even though some activities shown in section 3.3.1 can be executed simultaneously, it was decided to simulate the tasks sequentially. This was decided in order to simplify the execution. The indicators associated with each task is assumed to not be affected by other tasks who happen simultaneously, which allows for the sequential execution of the program.

3.4 Machine Learning

This section covers relevant aspects of the machine learning, including exploration of the datasets, preprocessing, train test split, selection of the models' hyperparameters and feature importance. It was decided to perform classification with the machine learning algorithms presented in section 2.5.2. These are decision trees, random forest, AdaBoost, XGBoost and SVM. All machine learning related programming was done in Python. Scikitlearn provided the algorithms except for XGBoost which originated from its own library.

3.4.1 Datasets

Two datasets were created. One dataset was created using an accident rate for serious accidents. This resulted in a fairly balanced dataset with a similar number of projects with accidents to the number of projects without accidents. The second dataset was created using an accident rate for fatal accidents. This caused a dataset with very few projects having fatal accidents. This is therefore considered an imbalanced dataset. As mentioned in section 3.1.3, the datasets consisted of only planning indicators and the target feature being the number of accidents. The construction indicators were removed in order to investigate whether planning data would be enough to predict future safety performance in the projects. This applied to both datasets; serious accident rate and fatal accident rate sets. An illustration of a dataset is shown in figure 31.

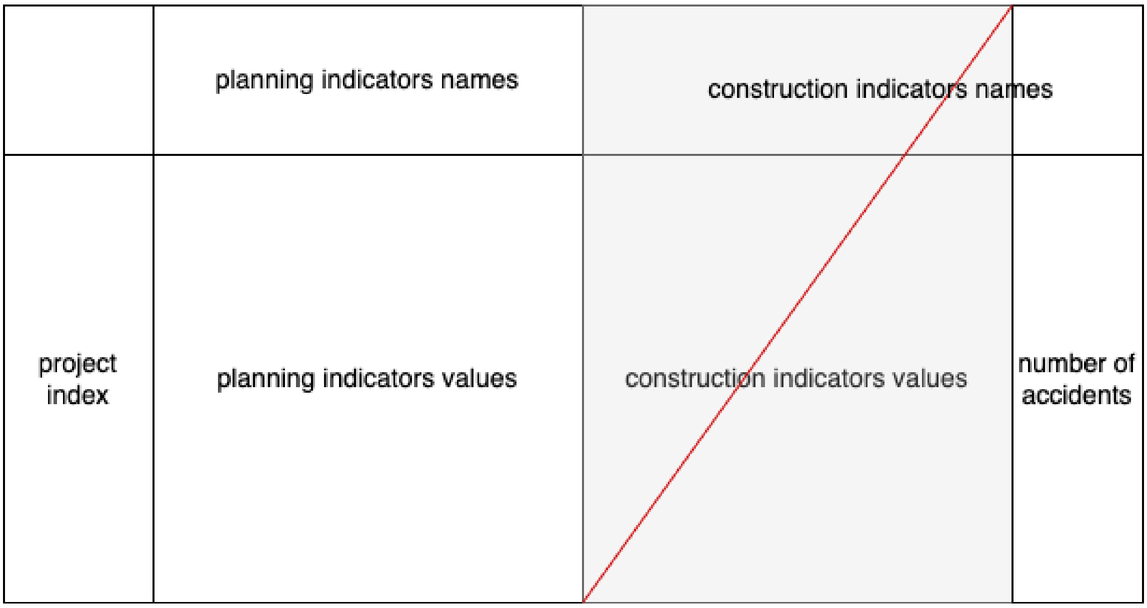


Figure 31: Dataset used for machine learning

Each dataset was generated with 10 000 projects, resulting in the same amount of rows. The number of columns reflects the number of indicators, 53 columns from the planning phase and one column from the construction phase with the number of accidents in the project. In figure 32 the balanced dataset is visualized. It is can be seen that around 58% of projects have one or more serious accidents.

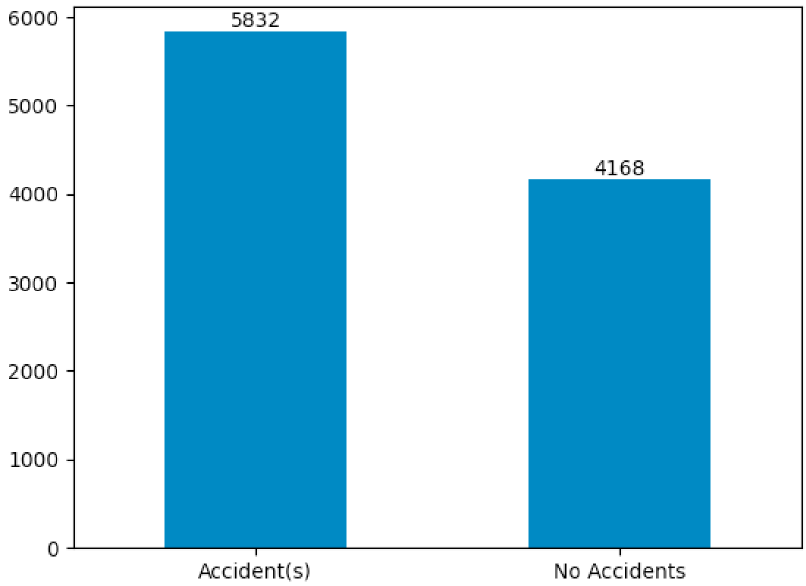


Figure 32: Amount of projects with no accidents and serious accidents

When visualising the imbalanced dataset the difference between the number of projects with a fatal accident and the number of projects without a fatal accident is clear. This visualisation can be seen in figure 33. Around 0.5% of all projects have a fatal accident.

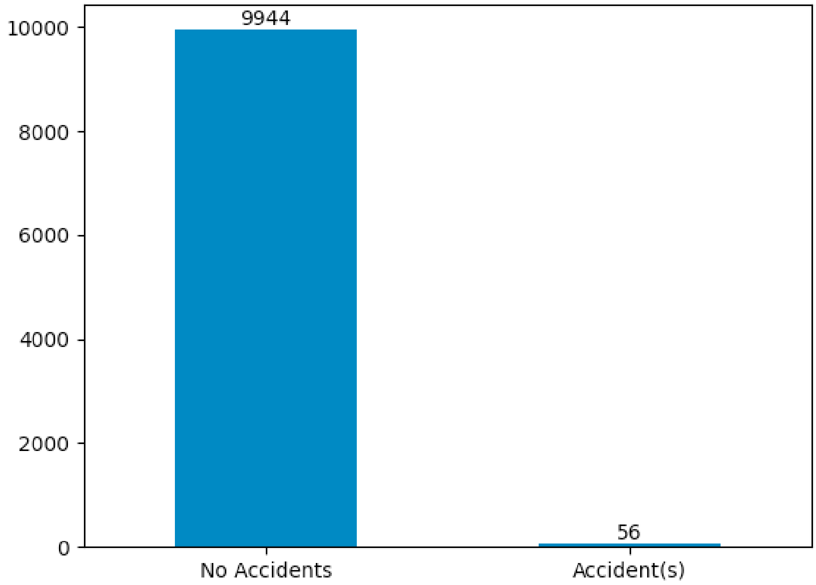


Figure 33: Amount of projects with no accidents and fatal accidents

The total project duration of all the projects is shown in figure 34. The different project durations have been grouped together. The most common durations are between 1000 and 1499 days, but durations between 500 and 999 days are also common.

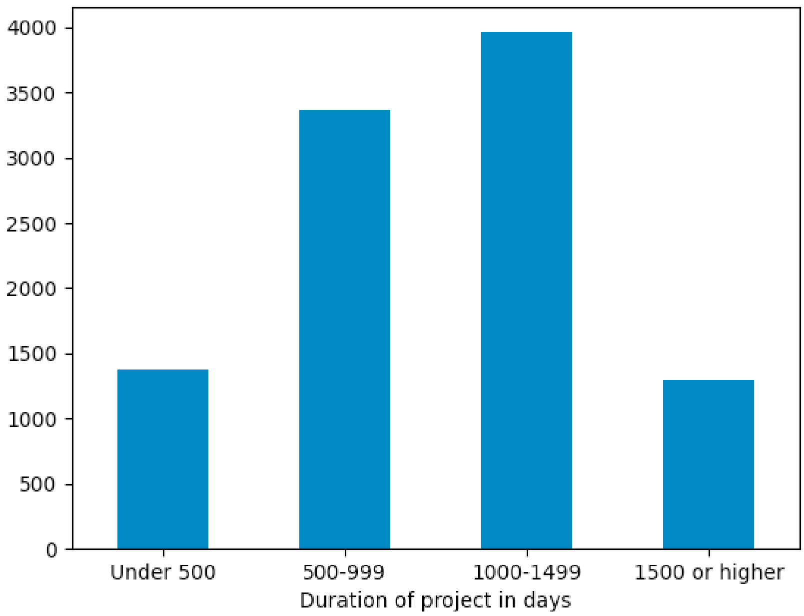


Figure 34: Amount of projects with different durations

The distribution of different budgets is shown in figure 35. Budgets in the same range have been grouped together. Most projects have a budget between 250 MNOK and 499 MNOK.

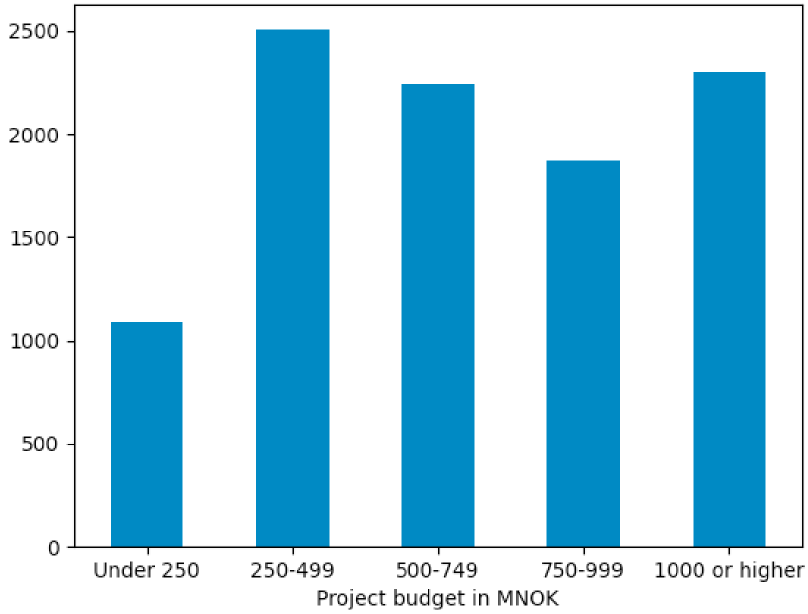


Figure 35: Amount of projects with different budgets

3.4.2 Preprocessing

As the datasets consisted of only integer values, they required a small amount of preprocessing. The target feature for both datasets was a column stating how many accidents had occurred in a project, while it was desired to perform binary classification. Hence, this column was transformed to true/false values, represented by the integers 1 and 0. All projects with one or more accidents were mapped to 1, while the projects with no accidents were mapped to 0.

3.4.3 Train Test Split

In order to train and test the models, the datasets were separated into two parts. A training set is necessary in order to train the models, while a test set is used to evaluate the models (Russell and Norvig, 2016). The final evaluation should be objective and the models should therefore not be trained on the same samples as it will be tested on. The split percentage was set to 75% for the training set and 25% for the test set. Shuffle was set to true in order to ensure randomness to which projects are in which sets. This would be important if only certain parts of the original dataset contained positive occurrences. Random state was also defined, in order to recreate the same train and test sets multiple times.

3.4.4 Selected Hyperparameters

Hyperparameters are parameters of the model class (Russell and Norvig, 2016). Selection of hyperparameters can be used to modify the machine learning models. Table 15 shows hyperparameter values which were used for the models in this thesis.

Model	Hyperparameter	Value
Decision tree	criterion	entropy
	splitter	best
	min_samples_split	2
	min_samples_leaf	1
Random forest	n_estimators	100
	criterion	entropy
	min_samples_split	2
	min_samples_leaf	1
	max_features	sqrt
AdaBoost	n_estimators	50
	learning_rate	1.0
	algorithm	SAMME.R
XGBoost	eta	0.3
	eval_metric	logloss
	max_depth	6
	sampling_method	uniform
SVM	kernel	linear
	degree	3

Table 15: Selected model hyperparameters

There are many possible hyperparameters for decision trees. *Criterion* is the function to measure the quality of a split. *Splitter* allows to choose between split strategies. *Min_samples_split* describe the minimum number of samples required to split an internal node. *Min_samples_leaf* describe the bare minimum of samples that must be present at a leaf node. For the random forest model, the number of estimators is also selected. *N_estimators* represent the number of trees in the forest. Additionally, *max_features* determine how many features should be considered when finding the best split. Selecting *sqrt* results in the max features to equal the square root of number of estimators. AdaBoost also has the hyperparameters *learning_rate* and *algorithm*. *Learning_rate* is the weight applied to each classifier at each boosting iteration, while *algorithm* is the boosting algorithm. For XGBoost, *eta* is the equivalent to *learning_rate*. *Eval_metric* determines the evaluation metric for validation data. *Max_depth* is the maximum depth of a tree, while *sampling_method* is how to sample the training instances. At last, some hyperparameters for the SVM model are *kernel* and *degree*. *Kernel* specifies the kernel type to be used in the model. *Degree* is the degree of the polynomial kernel function.

3.4.5 Feature Importance

Correlation heat maps were generated to explore correlations between the different features. A correlation describes a mutual relationship or connection between two features. The correlations towards the target feature may also indicate potential predictive patterns that the algorithms may look for. The correlations are represented through a matrix. The diagonal of the matrix has 1 as values, as each feature has a perfect positive correlation to itself. The correlation matrices for our datasets can be found in appendix A. As each dataset consists of 54 features in total, it was decided to separate them into several matrices. They were separated based on their planning stage, being either strategic definition, concept development, concept processing or detailed engineering. The target feature was included in all correlation matrices. This separation improved the readability, but also provided some limitations as several correlations were not shown between the stages.

First, correlations within the strategic definition stage were examined. This included indicators from the business case and the gap analysis activities. Generally, perfect positive correlations are found where the function of one indicator is directly connected to another. This occurred between several duration indicators. Further, high correlations were found between complexity/size and duration. *BusinessCase.RiskAssessment* had several strong negative correlations towards the features concerning duration, size and complexity. With regards to the target feature *numberOfSevereAccidents*, the features concerning duration, size and complexity had the strongest positive correlations. These were approximately 0.4. Further, negative correlations were found between *BusinessCase.Quality* and *BusinessCase.RiskAssessment* towards the number of accidents.

Secondly, the correlation matrix for the concept development stage was analyzed. It included indicators from the concept documents, financial framework and functionalities and sketches activities. Similarly to the strategic development correlation matrix, the features describing duration had strong positive correlations. In addition, the budget was highly correlated to duration. Further, the quality indicators were highly correlated to each other (0.97). The project status indicators were negatively correlated to the quality indicators. These correlations were around -0.4. It can be seen that the features having the highest correlations to the number of accidents were the ones concerning duration.

For the next planning stage, concept processing, correlations between indicators within the financial plan, regulation plan and preliminary project were examined. In addition to duration, high correlations were found between *FinancialPlan.EffectiveRiskManagement* and the quality features. Negative correlations were found between the project status features and the quality features. These ranged between -0.38 and -0.58. Even though the strongest correlations towards the target feature were still associated with duration, one can see that the project status features also had correlations of 0.11, 0.13 and 0.14.

Finally, the correlation matrix for the detailed engineering stage was reviewed. This matrix consisted of indicators from the plans for execution activity. All the indicators that were used as direct input to the construction phase model can be found here. *PlansExecution.LevelOfDetail* had a correlation of 0.62 towards*PlansExecution.Quality*. Another finding is that the schedule pressure had a correlation of 0.37 towards the number of accidents. Other features that had significant correlations towards the target feature were *PlansExecution.EquipmentLiability* (-0.1) and *PlansExecution.ProjectStatus* (0.12).

For the fatal accident rate dataset, the majority of correlations remained consistent with the ones for the serious accident rate dataset. The only difference was the correlations towards the target feature. Here, the values were significantly smaller. This was due to the implementation of the serious and fatal accident rates being in the construction phase model, rather than in the planning phase model.

3.5 Evaluation of the Method

Reliability and validity are two important concepts in research methodology that can be used to assess the quality of the method. These are presented below in the context of this thesis.

3.5.1 Reliability

Fellows and Liu state that "reliability concerns the consistency of a measure" (Fellows and Liu, 2003). For this thesis, reliability can be measured towards both the system dynamics and the machine learning. Reliability of the system dynamics model regards to what degree the model can be reproduced. The method provides a clear explanation of the technical implementation, activities, and indicators, making it highly reproducible. Still, some details regarding the extent of impact and uncertainties within each indicator is left out of the thesis. Reliability is also a measure of consistency of the datasets. Due to the implemented uncertainty, the datasets will have small variations when generated several times. Still, due to the big amount of generated projects, the dataset will hold very similar information between each run.

Consistency in regards to the machine learning is achieved by using seeds for the train-test-split and within the machine learning models. This way, the same results will be produced when running the machine learning models several times. Consistency of which projects belong to the train and test data also results in more reliable comparisons between the selected machine learning algorithms.

3.5.2 Validity

A measure's validity refers to how well it captures the concept it is intended to measure (Fellows and Liu, 2003). For this thesis, validity is based on determining if the findings are accurate from the project organizations' point of view. Qualitative validity in this context refers to the trustworthiness of the inferences that has been drawn from the theoretical background. Further, interviews with industrial partners could contribute positively to the validity of the method, but was not possible for this thesis. Creswell state that validity is one of the strengths of qualitative research (Creswell, 2009). For the quantitative research, the validity of the system dynamics model is a measure on whether one can draw meaningful inferences from the tests. The experiments done in section 4.1 shows validity on how the model has been developed. Still, comparisons to real data would increase the quantitative validity of the model as well.

4 Results

This section presents the results obtained by evaluating the system dynamics model and the machine learning models. The system dynamics model was evaluated based on extreme condition- and sensitivity tests, while the machine learning predictions were evaluated using several metrics. Both are tested for serious and fatal accidents.

4.1 System Dynamics Model

To validate the system dynamics model, extreme condition- and sensitivity testing was performed. Different values of the indicators were used as input from the planning phase to the construction phase. This was done through two experiments; first varying most indicator values which are on a range from one to ten, then varying the size and duration of the project. This separation of experiments was chosen because the correlation heat maps from section 3.4.5 showed a high correlation between project duration and the number of accidents. Hence, it was assumed that the impact from duration would overwrite the impact from the other indicators. Further, size is highly correlated to duration. Size and duration were therefore evaluated through a second experiment. Each experiment consisted of three different tests. For each of these tests, the simulation was run 1000 times with the same input indicators from planning. Then the target column, being number of accidents, was averaged. This was done twice per test; first for the dataset with a serious accident rate and secondly for the dataset with a fatal accident rate.

Experiment 1

The first experiment was to conduct extreme condition tests for the indicator values which were on a scale from one to ten. For this experiment, size and duration remained the same. The indicator values which were used are shown in table 16. Test 1 shows the worst case scenario, with most indicators holding the value 1. *SchedulePressure* holds the value 10, as a high pressure value expresses bad performance. Test 2 was done to explore the medium case scenario. The mean value of these indicators would be 5.5, but 5 was selected as the model only holds integer values. Test 3 tested the best case, and held indicator values of 10. Here, *SchedulePressure* was set as 1. For all three tests, size was set to 5 and duration as 500 for consistency.

The results show a trend where the number of accidents decreases as the indicator values improves. To interpret the results, one could say that the average number of serious accidents would be 1.338 by having the worst possible foundation from the planning phase. If a project has the best possible foundation from the planning phase, one would on average get 0.664 serious accidents in the project. The average number of fatal accidents is close to zero, however the trend is also found within the fatal accident experiment.

	Test 1	Test 2	Test 3
ProjectSize	5	5	5
EntreprenorCompetency	1	5	10
ConstructionDuration	500	500	500
PlansExecution.Quality	1	5	10
PlansExecution.SafetyTraining	1	5	10
PlansExecution.SafetySupervision	1	5	10
PlansExecution.SchedulePressure	10	5	1
PlansExecution.EquipmentLiability	1	5	10
numberOfSeriousAccidents	1.338	0.934	0.664
numberOfFatalAccidents	0.005	0.004	0.002

Table 16: Experiment 1 - indicator values

Experiment 2

The second experiment was conducted in order to see how much duration influences number of accidents. This shows how *sensitive* the model is to this specific indicator, making the experiment a sensitivity test. Size was also included in this experiment as duration and size are highly correlated in the model. The indicator values which were used are shown in table 17. Test 4 shows the shortest possible duration of the construction phase measured in days. This was calculated to be 91.25 with the project size being the smallest possible. Test 5 was done to explore medium long projects. A duration of 500 was chosen for this test, as it roughly represents the middle duration. Test 6 shows the longest possible construction duration for the model, being 912.5 days, with its associated size 10. All other indicator values were set as 5 for consistency.

Not surprisingly, the results show that there is a higher average number of serious accidents when the project goes over a longer time. For the fatal accidents, this trend is not as linear. Here the medium length projects and the longest projects have the same average number of fatal accidents.

	Test 4	Test 5	Test 6
ProjectSize	1	5	10
EntreprenorCompetency	5	5	5
ConstructionDuration	91.25	500	912.5
PlansExecution.Quality	5	5	5
PlansExecution.SafetyTraining	5	5	5
PlansExecution.SafetySupervision	5	5	5
PlansExecution.SchedulePressure	5	5	5
PlansExecution.EquipmentLiability	5	5	5
numberOfSeriousAccidents	0.162	0.927	1.773
numberOfFatalAccidents	0.001	0.006	0.006

Table 17: Experiment 2 - indicator values

These two experiments contribute to the validation of the model. The results which were obtained provide evidence of the model's reasonableness. As this thesis contains the planning phase only, the trends were examined, rather than the actual accident numbers as this was implemented in the construction phase model.

4.2 Machine Learning

The machine learning predictions were evaluated based on four metrics; accuracy, precision, recall and F1-score. These metrics were described in section 2.5.3. Confusion matrices were also generated for all models in order to visualize classification patterns. This was done for both the balanced dataset consisting serious accidents and the imbalanced dataset consisting fatal accidents.

4.2.1 Serious Accidents

Table 18 shows the results of the five machine learning models which were tested on the serious accident dataset. SVM and AdaBoost obtained the best results overall. AdaBoost had the best accuracy and precision, while SVM had the highest recall. Their F1-scores were equal and higher than the other models'. The decision tree model performed the worst according to all the selected evaluation metrics.

Model	Accuracy	Precision	Recall	F1
SVM	0.677	0.670	0.861	0.753
XGBoost	0.639	0.661	0.757	0.706
AdaBoost	0.687	0.685	0.836	0.753
RF	0.671	0.676	0.815	0.739
DT	0.572	0.625	0.628	0.626

Table 18: Classification results with serious accidents

All models obtained accuracies below 70%. These values are considered low. Further, all precision values were also below 70%. The recall values were significantly higher, with SVM deriving 86%. The F1-scores were found between each model's precision and recall values. The confusion matrices below show what predictions these values are based on (figure 36, 37, 38, 39 and 40).

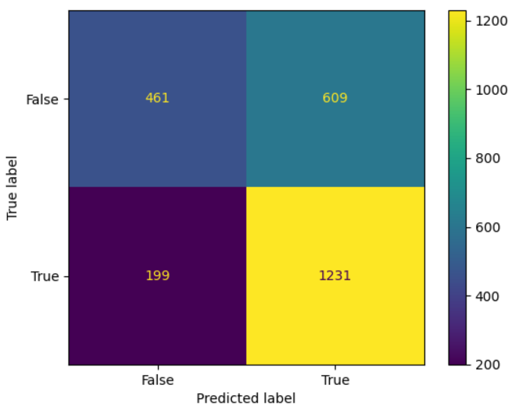


Figure 36: SVM confusion matrix

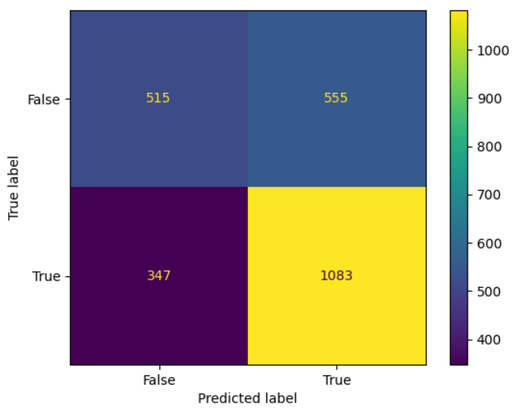


Figure 37: XGBoost confusion matrix

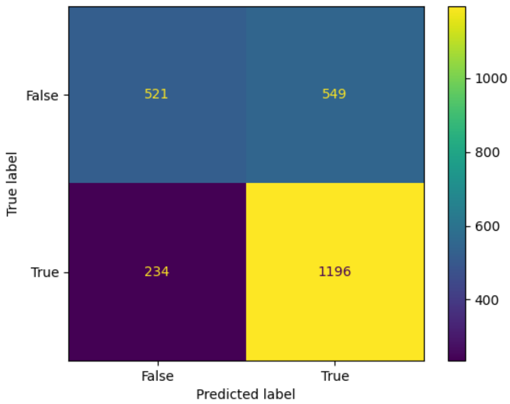


Figure 38: AdaBoost confusion matrix

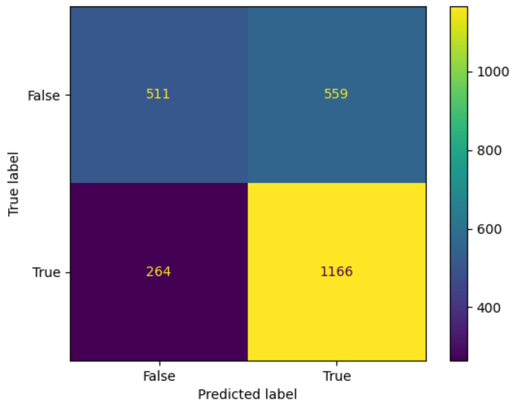


Figure 39: Random forest confusion matrix

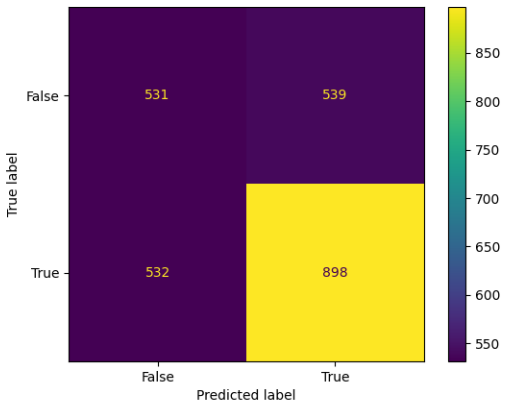


Figure 40: Decision tree confusion matrix

According to the confusion matrices, the models are generally better at predicting the positive instances than the negatives. For the negative instances are approximately half of the samples classified correctly. SVM predicts less true negatives than the others (figure 36). It generally predicts more samples as positives, which also results in more true positives than the other models. The decision tree model predicts less positive instances correctly than the rest (figure 40).

4.2.2 Fatal Accidents

Table 19 shows the results of the five models which were tested on the fatal accident dataset. The accuracies were very high while the other metrics were zero. This is also shown in the confusion matrices (figure 41, 42, 43, 44 and 45), where almost all samples are true negatives. None of the machine learning models were able to predict true positives, resulting in precision, recall and F1 values of zero. Even though the accuracies were high, the results are considered to have low values based on the other evaluation metric results.

Model	Accuracy	Precision	Recall	F1
SVM	0.995	0	0	0
XGBoost	0.995	0	0	0
AdaBoost	0.994	0	0	0
RF	0.995	0	0	0
DT	0.988	0	0	0

Table 19: Classification results with fatal accidents

The confusion matrices show that none of the 13 positives were classified correctly. The SVM, XGBoost and random forest models predicted no samples as positive (figure 41, 42 and 44). AdaBoost predicted one sample as positive (figure 43), while the decision tree model predicted 18 as positive (figure 45). All of the positive predictions were incorrect.

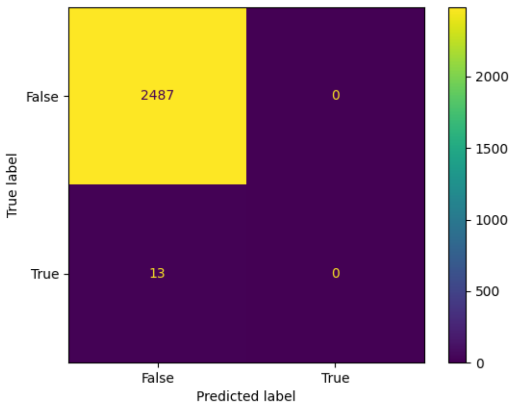


Figure 41: SVM confusion matrix

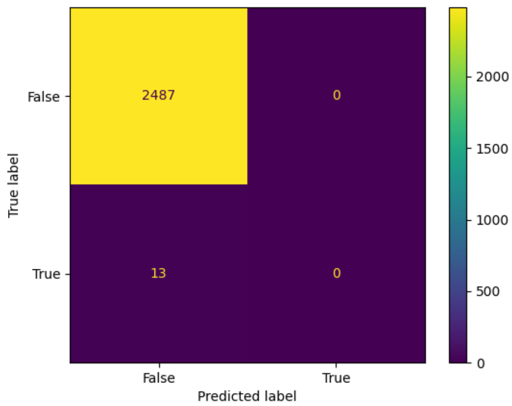


Figure 42: XGBoost confusion matrix

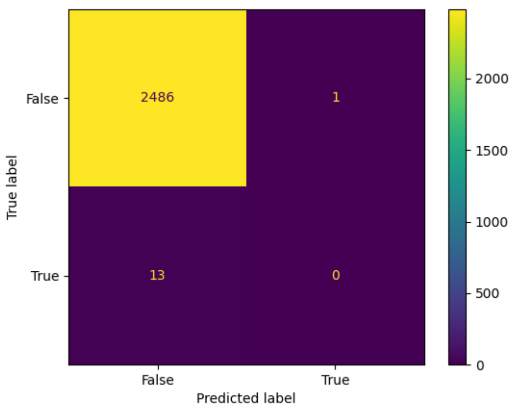


Figure 43: AdaBoost confusion matrix

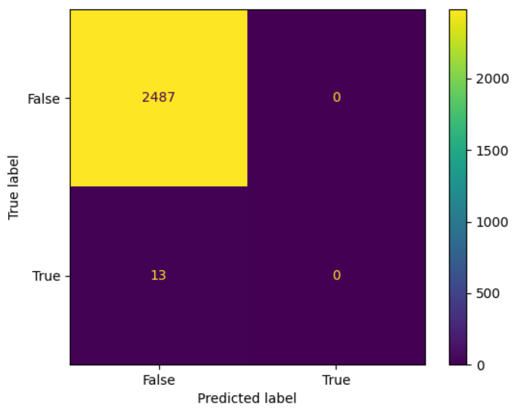


Figure 44: Random Forest confusion matrix

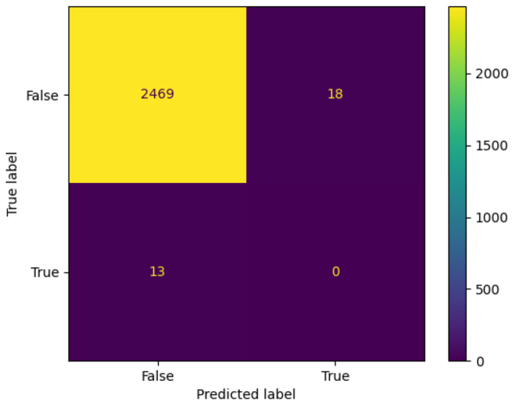


Figure 45: Decision tree confusion matrix

5 Discussion

This section provides a discussion on all the conducted work. This includes the development and testing of the system dynamics model and the machine learning. Additionally, their applicability within construction projects and combination of technologies are discussed. The section debates the methodology and results in connection to the objective of the thesis and the theoretical findings.

5.1 System Dynamics

A system dynamics model was developed to simulate construction projects and generate data for the machine learning. Further, the model was validated. The development and validation is discussed below. This covers only the planning phase, while discussions regarding the construction phase can be found in *System Dynamics Modelling and Machine Learning to Improve Safety in Construction Projects* (Aamlid, 2023).

5.1.1 Structure of the Model

As described in section 3.3.3, the model was built with projects, tasks and indicators. Each project had several tasks which adjusted the indicators during the project. The tasks represented the activities from section 3.3.1, being *Business Case*, *Gap Analysis*, *Concept Documents*, *Financial Framework*, *Functionalities and Sketches*, *Regulation Plan*, *Financial Plan*, *Preliminary Project* and *Plans for Execution*. These were based on the planning stages and deliveries presented in the theory (section 2.1.3). Other structures could be considered as well, as it is hard to define exact activities. A potential risk is leaving out certain processes to the model. Japkowicz and Shah considered oversimplifying as a threat of using artificial data (section 2.5.1). Still, construction projects are very complex (section 2.1.2) and simplifications have to be made when developing any system dynamics model. This is in accordance to Fellows and Liu who stated that models should reflect the reality as closely as is practical, but be reasonably cheap to construct and use (section 2.3.1).

Further, it was decided to execute the activities sequentially. Based on the overall structure including stages and activities (section 3.3.1) it can be seen that some activities could be executed in parallel. This applies to the activities within the same planning stage. This is due to the input coming from a previous stage. In the theory, Hansen stated that the early phase planning consisted of reciprocal dependencies, while the detailed engineering and building phase is often more sequential (Hansen, 2019). Still, it was decided to execute the activities sequentially in order to simplify the implementation. The sequential execution does however limit the possibility of several activities affecting each other in loops. Hence, there are no *feedback loops* in between the activities (section 2.3.1). It can be argued that in the real world certain activities might affect each other in a back-and-forth manner. Consequently, executing the activities in parallel could result in a more accurate representation. However, the sequential structure made it easier to create separate influencing diagrams and put focus on the interactions between indicators within each activity (section 3.3.2). Additionally, as mentioned above, the model should be reasonably cheap to develop and use (Fellows and Liu, 2003). Thus, sequential execution was considered the most appropriate for this thesis.

The class based implementation of the model allows for easy extension and modification for further work (section 3.3.3). As mentioned in the methodology section, it was decided to focus on one type of project (section 3.3). In the future, one could add other types of projects with only small modifications to the program. This could be done by adding or changing tasks representing the activities. However, this applies more to the construction phase which is not in the scope of this thesis. The planning phase has less variations regarding the activities, but the implementation has left room for easy modification there as well. Another perspective is the ability to add or change the chosen indicators for the planning stage. The chosen class structure facilitates for this. Hence, the choices done in regards to implementation provides for expansion and adjustments, which makes it more suitable for further work. Still, when utilizing the model to provide a dataset, different sets of indicators between the projects would result in missing values. As it was planned to use the dataset for machine learning, no missing values and therefore one type of project was an advantage.

5.1.2 Indicators

Appropriate indicators needed to be selected in order to simulate the different activities. The chosen indicators were shown in section 3.3.2. These indicators were chosen based on project- and safety management theory (section 2.1 and 2.2) and indicators used in previous system dynamics models of safety performance (section 2.3.3).

The literature review on system dynamics and safety performance showed some common indicators. These indicators were presented in table 1. Pressure was frequently used in other models. It was also presented as a safety factor in section 2.2.2. For this thesis, the model was given the *PlansExecution.SchedulePressure* indicator. Schedule pressure was chosen over productivity/production pressure since the model is still in the planning phase. Li et al. also included cost and schedule in their model. *FinancialFramework.Budget* and *ProjectSchedule* were selected for this thesis. In the model by Li et al. cost and schedule affected productivity pressure (Li *et al.*, 2022). Figure 29 shows how *SchedulePressure* was affected by *FinancialFramework.Budget* as well as the schedule related indicators *ProjectSchedule*, *ConstructionDuration*, *PlanningDuration* and *ProjectStatus*. While the assumption is that being behind schedule leads to increased pressure, it is not necessarily the case that being behind schedule in the planning phase will result in pressure during the construction phase. It is also assumed that if the project has a high budget relative to duration, there will be more pressure. This is due to a high budget being associated with a big amount of work to be done.

Supervision was the indicator that was included in the most system dynamics models (section 2.3.3). This was also a managerial factor discovered in the safety factors review (section 2.2.2). Therefore, *PlansExecution.SafetySupervision* was included as an indicator, being affected by both *EntreprenørCompetency* and *FinancialPlan.SafetyBudget*. Safety related indicators were used in many of the models. One indicator that was commonly used was safety training. Another safety related indicator was risk assessment. The model was therefore given the indicators *PlansExecution.SafetyTraining* and *BusinessCase.RiskAssessment*. Both of these were also listed as influential factors in section 2.2.2. The indicator *SafetyTraining* is affected by the same indicators as *SafetySupervision* and they will therefore be quite similar. To improve the complexity of the model these indicators could be affected by different indicators. Next, equipment was an environmental factor uncovered in the safety factors review. It was decided to include this in the model through the *PlansExecution.EquipmentLiability* indicator. The indicator was additionally based upon equipment condition from Sun et al.'s model. *PlansExecution.EquipmentLiability* is randomly set as a number between one and ten in the late stages of the planning phase. For a less random model this value could be based upon some of the other indicators.

Competency was represented in the model by both *ByggherreCompetency* and *EntreprenørCompetency*, describing the competency of the owner and contractor respectively. Competency was highlighted by numerous articles in the safety factors review (section 2.2.2). The competency of the owner is important as they are responsible for overall decisions, defining the project and choosing contractors (section 2.1.2). The competency of the contractor influences the project as they are the executing parties. Both competency levels were set at random, but further it could be looked into if the competency of the owner is connected to the competency of the contractor. More competent owners could be more likely to choose competent contractors. The SHA plan is established during the preliminary project and is therefore formed by the owner. Thus it is assumed that the competency of the owner combined with the quality of the concept documents affects the quality of the SHA plan. Although the indicator *PreliminaryProject.SHAPlanQuality* was created during the model development process, it is not utilized further. This could be improved by either having the SHA plan affect the quality of the preliminary project or having it as one of the indicators being used further in the construction phase. The quality of the preliminary project itself was set to affect the level of detail in the plans for execution. It was considered more probable to develop a detailed plan for execution if the relevant previous work was done thoroughly.

The quality of each activity was deemed to be important, as the quality of a project could minimize the risk of an accident (section 2.1.2). Therefore each activity, except for the regulation plan, was given a quality indicator. Quality was not considered an appropriate measure for the regulation plan, since the application is either approved or declined. Since the regulation plan is formed by the owner and approved by the local authorities (section 2.1.3), coordination between these two parties is required. This resulted in establishing the indicator *RegulationPlan.CoordinationAuthorities*, which is used to show that poor coordination between local authorities and the owner could lead to delays. Other factors which were assumed leading to delays were complexity and stakeholder requirements. As presented in section 2.1.2 projects with a high level of complexity would require more complex and varied methods, practises and processes. It was therefore assumed that a longer duration is needed to finish the project. The complexity of the project was also set to affect the size of the project in terms of cost, which is considered reasonable due to the requirements mentioned above. Further, it was assumed that complexity affects the quality of the risk assessment. This is due to that when there are more factors to consider, high quality risk assessment will be more difficult to create. Next, stakeholder requirements are considered a risk in terms of delays. Section 2.1.3 states that the gap analysis measures the gap between the current situation and the users needs. If there are more requirements, there are more aspects to take into consideration during this stage, which might take a longer time to process. Therefore the indicator *GapAnalysis.StakeholderReq* is set to affect the actual duration of the gap analysis activity.

Some indicators were identified in the theory, but not utilized for this model. Many of the studies that were reviewed had indicators such as safety climate/culture and management's commitment to safety. This could have been added to the model to simulate the safety related factors of the project with even more complexity. However, they were not included in order to simplify the model. One could potentially argue that *FinancialPlan.SafetyBudget* represented the management's commitment to safety, as a large safety budget would mean the management prioritizes safety. Still, this could be implemented as a separate indicator. Further, some indicators were present in multiple of the system dynamics models from the literature review, but were still not included in this model. Mohammadi and Tavakolan included indicators such as fatigue, rework and incident investigation in their model (Mohammadi and Tavakolan, 2019). Jiang, Fang and Zhang also had incident learning and Li et al. had rework in their model (Jiang, Fang and Zhang, 2015; Li *et al.*, 2022). These indicators were not included because they were deemed to be in the construction phase and therefore not relevant for the planning phase model.

When looking at what indicators have a high correlation with the number of accidents there are some unexpected results. As described in section 3.4.5 all the duration indicators have a high level of correlation with the number of accidents. This is considered reasonable since longer projects are assumed more likely to experience accidents than short projects. Further it can be seen in the correlation matrices that *PlansExecution.SchedulePressure* has a relatively high correlation with the number of serious accidents being 0.37 (appendix A). This is unlike *PlansExecution.SafetyTraining* and *PlansExecution.SafetySupervision* that only have correlations of -0.029 and -0.03. This differs from the many other studies that had safety training and supervision as indicators closely connected with the accident rates in their system dynamics models. This discrepancy should be investigated further, and efforts should be made to establish reasonably high correlations between all indicators passed on to the construction part of the simulation and the number of accidents. Lastly, many of the functions have a degree of randomness in them. This is to account for randomness in the real world, but the real life level of randomness is unknown.

5.1.3 Validation of the Model

Extreme condition tests were conducted in order to validate the planning model (section 4.1). The results of *Experiment 1* showed a trend where poor indicator values from the planning phase resulted in more accidents and opposite. This seems reasonable. The average number of serious accidents for the worst values was approximately twice the number as for the best values. Medium indicator values gave an average accident number which was closer to the good values than the poor. This implicates a non-linear relationship between planning indicators and safety performance in the model. This might reflect the real world, where many projects have decent safety performance, while some projects with really bad safety performance result in many accidents. The trend for the fatal accidents was similar, but not of the same proportions as for the serious accidents. This is probably due to the low number of fatal accidents. Fewer values give more uncertainty as there are less values to draw the average from.

The sensitivity tests for duration (*Experiment 2*) showed bigger differences in the number of accidents than for the first experiment (section 4.1). It is obvious how longer projects will have more accidents than shorter projects on average. This is also shown in the correlation matrices in section 3.4.5. This could be an argument to choosing another safety measurement or potentially combining several different measurements as suggested by Kjellen and Albrechtsen (section 2.2.1). Hence, this experiment shows why it could be beneficial to measure the safety performance as a rate. In regards to the average number of accidents, that is determined by a number of contributing factors implemented in the construction phase (Aamlid, 2023). The actual number of accidents is therefore not taken into account for the experiments and will not be discussed in this thesis, except for the trend influenced by the planning phase.

The literature review in section 2.3.3 proposes several ways to validate system dynamics models. Qayoom and HW Hadikusumo used extreme condition testing such as in this thesis (Qayoom and HW Hadikusumo, 2019). Other studies suggested sensitivity testing or comparing the simulations to real data. Sensitivity testing was performed in accordance to the theory. However, this was for the duration and associated size indicator only. This could also be considered for other indicators in order to evaluate how they influence the number of accidents. Still, correlations between single indicators and the number of accidents were also explored within feature importance (section 3.4.5).

Real data could also be used for validation and/or calibration of the model. Comparing the simulations with real data was a common approach according to the literature review on system dynamics and safety performance (section 2.3.3). Additionally, Nikolenko stated that a common approach is training the model on synthetic data with the intention of using it on real data (Nikolenko, 2021). For this research, real data was difficult to gain access to and collect. Ideally, one would have access to project data representing all the indicators and therefore being able to explore and verify their influence on each other. There are several difficulties regarding this; a lot of this data is highly qualitative and one would have to define ways to transform them into quantitative data. Secondly, some data is confidential and can not be accessed due to privacy reasons. Examples include the competency of the people involved in the construction project. Nikolenko stated that this is also considered a reason for using synthetic data (section 2.5.1). Lastly, one would need data from both the project owner and the contractor to cover a project from start to end. This data gathering requires a lot of time and collaboration between the parties. Hence, calibration of the model would be useful to make it reflect the real world, but was not possible for this thesis. Calibration based on feedback from industrial partners was also considered as an option, but no feedback was available. Calibration of the model could however be an option for further work. A realistic approach would be to start calibrating the parts of the model where there is available data or put the focus on the parts which are the most uncertain.

5.2 Machine Learning

Several decisions were taken when creating machine learning models, including selecting classification, machine learning algorithms, and hyperparameters. These choices and their effects are discussed below. The results obtained by the machine learning models for both serious and fatal accidents are also discussed.

5.2.1 Building the Machine Learning Models

For the machine learning, it was decided to perform classification. As the number of accidents column from the simulation model contained continuous values, one could by definition perform regression as well (section 2.5). However, classification was considered a more appropriate learning problem due to two reasons. Firstly, the results from classification are easier to interpret. For this thesis, it would be easier to evaluate how many projects were classified correctly than to evaluate how far off the machine learning predictions were in terms of accidents within each project. Secondly, it was considered more important to predict which projects had the biggest risks of accidents than predicting the exact number of accidents. This was due to the objective of the thesis; giving early warnings of projects with a high accident risk. Additionally, there is a lot of uncertainty which makes the latter less realistic.

Five algorithms were selected to be used in this thesis; decision trees, random forest, AdaBoost, XGBoost and SVM. This selection was based on both the literature review on machine learning for safety predictions (section 2.5.4) and the literature review on combining system dynamics and machine learning (section 3.1.3). The decision tree algorithm was widely used in the literature review on machine learning for safety predictions. It was also used to support the system dynamics method in the paper by Duggan (Duggan, 2020). Random forest was even more widely used in the first literature review and also outperformed all other models in several of the articles. It was also combined with system dynamics by Edali and Yücel (Edali and Yücel, 2020). XGBoost was not as widely used, but did outperform the other algorithms in all the studies it was used for (section 2.5.4). Additionally, XGBoost is a popular algorithm for tabular data (section 2.5.2). It was therefore considered a good candidate for this thesis. AdaBoost was another boosting algorithm which was tested. It was only used in two of the other studies. SVM was also widely used and stands out to the other tree based algorithms. It's regression version SVR was also used together with MLP in a previous research by Roozkhosh with data from a system dynamics model (section 3.1.3). One algorithm type which was widely used within the literature was variations of neural networks. One could have tested this instead of the AdaBoost algorithm, however it was decided to not test any more algorithms due to the low performance.

Hyperparameters which were used for this thesis are shown in table 15. Selection of hyperparameter values can be used to modify the machine learning algorithms and potentially increase their performance slightly. Due to the low performance it was decided to not test several configurations. To increase the performance, attention should rather be put on the dataset itself. If one were to achieve good results in the future, one could consider testing several hyperparameters in order to find the optimal model for this purpose.

5.2.2 Results

The machine learning predictions for serious accidents resulted in accuracies among 60-70% and recall values around 80% (section 4.2.1). These results are considered unsatisfactory. The highest accuracy and precision were achieved by AdaBoost. This is an interesting aspect as AdaBoost was not included in many of the previous machine learning for safety predictions studies (section 2.5.4). Despite SVM obtaining less true negatives, the model provided the best recall value of 86%. Both of these models gave equal F1 values. The decision tree model gave the lowest values, including a recall of 62.8%. This is reflected through significantly less true positives than the other models, as seen in the confusion matrices. The low values can be a result of too small correlations between the planning phase indicators and the construction phase indicators. As presented in section 3.3.2, only eight planning indicators were given as input to the construction model. In the correlation matrices it is evident that only *PlansExecution.SchedulePressure*, *ProjectSchedule* and *ConstructionDuration* out of these eight have significant correlations to the number of accidents. *PlansExecution.EquipmentLiability* also show some correlation to the number of accidents, while the rest are very insignificant. For further work one could either include more of the planning indicators as input to the construction indicators or adjust the functions for the construction indicators such that the planning indicators have a greater importance. Better results could also possibly be achieved by calibrating the model, which was discussed in section 5.1.3.

Despite the low values, the recall value of 86% shows an important perspective. As presented, recall measures how many of the actual positives that are predicted correctly (section 2.5.3). The actual positives in this thesis are projects with accidents. Since one is specifically interested in the positive instances, recall is a fitting measurement. With greater recall values, lower precision values are tolerable for this type of thesis. This is because even though the machine learning model predicts more values as positives which are actually negative, the most important aspect would be to separate those who have a higher probability for accidents. This is similar to within the healthcare industry, where it is considered better to have false positives than false negatives. As seen in the confusion matrices, the machine learning models are better at predicting the positive than the negative instances (section 4.2.1). Still, the recall value does not capture all aspects of the predictions, and should be combined with other metrics. This is in accordance to Japkowicz and Shah (Japkowicz and Shah, 2011). Overall, despite the low values a potential is seen in the higher recall values and even better measurement values seem in reach for further work.

For the fatal accidents, the results are very different than the serious accidents. The accuracies are extremely high, while the other metrics have values equal to zero (section 4.2.2). This implies a very imbalanced dataset, which is true for the fatal accidents. As seen in the confusion matrices, several of the machine learning models predict all values to be negative. This gives the high accuracy values, as very few of the samples are actually positive. Both AdaBoost and the decision tree model predicts some positive instances, but they are all actual negatives. Hence, there are no true positives, resulting in precision, recall and F1 as zero. This shows the difficulties of predicting fatal accidents. There is insufficient data to learn from as there are very few accidents. There is also much uncertainty, which might imply that one should put the focus on predicting bad safety performance in general and not whether a fatal accident will occur or not. This is in accordance with the correlation matrices that also show very low values from the planning indicators towards the number of fatal accidents (section 4.2.2). One action that could be considered is trying different techniques to balance the dataset, as presented by Jo (section 2.5.1). Still, it can be argued how prediction of fatal accidents is not the most realistic approach. As seen in the literature review on machine learning for safety predictions (section 2.5.4), the objective of such research is often either prediction of whether an accident will occur or prediction of the outcome. For further work, one could predict the probability of accidents or other safety measurements, instead of predicting whether a certain type of accident will happen. Potential safety measurements include loss-based indicators such as the TRI rate or process-based indicators such as percentage compliance with regulations or number of RUO.

5.3 Applicability

The application of system dynamics and machine learning is discussed below. The first subsection covers combining these two technologies, whereas the second subsection focuses on how this can be utilized and bring value to the construction industry.

5.3.1 Combining System Dynamics and Machine Learning

In this thesis, a system dynamics model was used to create two synthetic datasets. These datasets were then used as input for training multiple machine learning models. As described in section 2.5.1 using simulated data as input for machine learning is nothing new. Many problems of AI are caused by limited data and synthetic data is a way of handling this. It enables the use of variables that are possible but lack data (Japkowicz and Shah, 2011). Still, one of the dangers with synthetic data is the risk of oversimplifying problems (section 2.6).

This approach of utilizing system dynamics to generate data that is used as input for machine learning has been done in other studies previously. Chen et al. used system dynamics to model a greenhouse and then machine learning to predict temperature and humidity based on the data from the system dynamics model (section 2.6). Roozkhosh et al. also used data from their system dynamics model as input for machine learning that further analysed the data (section 2.6). The disadvantages and inherent biases of using synthetic data for machine learning should be recognized. The quality of the system dynamics model has a significant impact on the precision and reliability of the synthetic dataset. The assumptions made when creating the model will directly impact the data collected from the model. It is important to take precautions to ensure that the synthetic data accurately depicts the nuances and complexity of the real-world system.

Other studies have used machine learning as a way of creating and improving the system dynamics model. An example is Abdelbari and Shafi who looked into using computation methods to create a system dynamics model directly from data (section 2.6). This way of using machine learning to improve the system dynamics model is a different approach of combining system dynamics and machine learning. Here, the focus is on enhancing the system dynamics model, whereas in other studies system dynamics is used first as a tool to aid or enable machine learning later. As mentioned in section 2.6 a typical approach is using synthetic data to train a machine learning model with the aim of using the model on real data in the future (Nikolenko, 2021). The machine learning models are currently implemented in a way that would make it easy to use them on real data if this was to become available in the future.

5.3.2 Applicability Within Construction Projects

The system dynamics model can be used to spread knowledge about safety in construction projects by demonstrating how various factors impact safety performance. This could be a valuable tool for various individuals, such as project managers or other personnel within organizations who have opportunities to influence safety management. The model could be used by itself to run the entire project simulation or in combination with machine learning to focus on a certain parts of the project. One option is modifying the model into a game in order to play with different values and try to avoid any accidents. Another option is making a visual live representation of the model to be utilized in presentations and other settings. For this to be trustworthy, calibration of the model is required, and should be considered for further work.

To integrate the technology into ongoing projects, one could train a machine learning model on historical data and insert real values from the project to predict safety performance. This way, the safety predictions would serve as a *leading indicator* as defined by Kjellen and Albrechtsen in section 2.2.1. It could be integrated into a dashboard that gave a warning if the project was indicating bad safety performance. Ideally the indicator that affects the bad safety performance the most could be exposed such that changes could be made within the process. The technology as it is implemented now does not support uncovering which factor contributes the most negatively in each project, but this could be a potential for further work. This way, actions could be taken on the bad delivery or process so that a new and better value is inserted and the performance indicator changes towards the better. This would serve as decision support during the project and potentially increase the safety performance in projects. A challenge is how to measure the different deliveries. For this technology to be effective, it would require automation of the transformation from documents into indicator values. This requires more research and can be time consuming. A potential is starting with the indicators that are proven to affect the safety performance the most, and then expand the automation as new methods are developed.

For the technology to give more accurate predictions, one might have to include data further out in the projects. This means including indicators from not only the planning phase, but also the construction phase. The safety performance indicator could still predict future developments and therefore serve as a leading indicator (section 2.2.1). With this perspective, the dashboard could be used during the construction phase, providing warnings if the project has a high accident risk. Including the construction phase data could result in more accurate predictions, as there would be more information available. There is also more research on machine learning for safety predictions during the construction phase (2.5.4). Hence there is more previous research to rely on in this case, which could also lead to higher performance and bring more value into the projects.

6 Conclusion

The objective of the thesis was to utilize system dynamics and machine learning to give an early warning of projects with a high accident risk. For this research, a system dynamics model was developed in order to simulate construction projects and generate datasets. Machine learning models were trained and tested on these datasets in order to predict safety performance. The conclusion is separated into three subsections; main findings, contributions and further work.

6.1 Main Findings

The first research question concerns how system dynamics and machine learning can be combined to predict safety performance in construction projects. Consequently, a system dynamics model was developed in order to simulate construction projects. It consisted of activities and indicators that were based on theoretical findings. The model simulates one type of project, but was implemented for easy extension and modification. Extreme condition- and sensitivity tests were performed and showed reasonable trends between the planning indicators towards the number of accidents. The system dynamics model was successful in producing datasets that served as the basis for machine learning. Both serious and fatal accident rates were used, producing two separate datasets. Further, the datasets were utilized to accomplish the training and testing of several machine learning models. This was in accordance with the theory, where it was found that using synthetic data for machine learning purposes is a common method when having insufficient data. Furthermore, a literature review on system dynamics and machine learning showed that other studies had integrated these two technologies in a similar matter. The machine learning predictions formed a leading indicator which predicts the future safety performance in a project. The conducted approach has limitations regarding the system dynamics model as it was not calibrated towards real values. Still, it can be concluded that combining system dynamics and machine learning offers a promising approach to predicting safety performance, particularly when real project data is not available.

The second research question regards the safety management implications of utilizing system dynamics and machine learning in construction projects. For the serious accidents dataset, while some measurement values could be improved, there were notable positive aspects. In particular, the recall rate of 86% demonstrated significant potential for identifying accident-prone projects. It was discussed how lower precision values were tolerable together with higher recall values for this domain. This is due to the objective of the thesis; giving early warnings for the projects with a higher accident risk, being the true positives for binary classification. Further, the results of the fatal accident predictions had low values due to the few occurrences of fatal accidents. This implicates that other measurements could be more appropriate for assessing safety performance. Examples are using a probability or rate, instead of predicting whether a certain type of accident will happen. Moreover, the results imply that data from the planning phase might need to be combined with additional data from later stages in the project to improve the predictions. To conclude, the results implicate that predicting fatal accidents is less realistic due to the few occurrences. Still, the results show a potential of separating the projects with a higher risk of serious accidents and therefore giving an early warning to improve safety performance within construction projects.

6.2 Contributions

This thesis offers several contributions to theory. It provides a detailed system dynamics model for the planning phase of construction projects, based on comprehensive theory and literature reviews. This model provides understanding of the complex dynamics inherent in construction project planning. Secondly, the thesis provides evidence that system dynamics and machine learning are effectively combined for simulations and predictions. This demonstrates the potential for leveraging these technologies for future research and other practical implications. Moreover, the thesis proves that accident type is not an appropriate measurement for safety performance predictions, and that other measurements should be considered. It also showcases how planning data might need to be combined with construction phase data in order to obtain more reliable forecasts.

In terms of practical contributions, the system dynamics model and machine learning predictions have potential to provide decision support towards construction projects. The combination of technologies helps share knowledge about safety in construction projects by demonstrating how different factors affect safety performance. Further, it has been discussed how the implemented system dynamics model in combination with machine learning can be applied within construction projects. This could be implemented as a dashboard to serve as a leading indicator for providing decision support through the projects. However, to achieve this, the thesis acknowledges the need to form different project documents into quantitative indicator values.

6.3 Further Work

There are two main perspectives which can be considered for further work; calibrating the model with real data and extending the dataset for machine learning using parts of the construction phase. Calibration of the model has been discussed in section 5.1.3. This could be done in order to make the model more reflective of the real world and therefore producing more accurate data. However, this type of data for an entire project is difficult to obtain. Hence, for further work it is suggested to start with selected parts of the model where real data is available. In order to calibrate the model one would have to collect real data, look for patterns in the data and then adjust the indicators and how they interact in the model. Another option would be to conduct interviews with industrial partners to obtain feedback for the model.

The second aspect which could be considered for further work is including a subset of the construction phase data into the machine learning. This was discussed in section 5.3.2. For this thesis, the construction phase data was removed so that the features only consisted in planning indicators. As several of the evaluation metrics gave relatively low values, one could experiment with how far into the construction phase one would have to include data from in order to get good results. This would imply how far into the project one would typically be in order to predict future safety performance, as the planning phase might be too uncertain. Despite including construction phase data, the predictions would still serve as a leading indicator and could bring great value into the project.

In addition to these two propositions, one could consider if more planning indicators should be given to the construction phase part of the model. Also, the model could be tested more in detail using sensitivity testing. Consequently, one would be able to examine the singular impact of each indicator. For the machine learning, other algorithms and hyperparameters could be considered in the future, if one were to obtain higher prediction values by adjusting the system dynamics model.

References

Aamlid, J. S. (2023). *Integrating System Dynamics Modelling and Machine Learning to Improve Safety in Construction Projects.*

Abas, N. *et al.* (2020). 'Factors affecting safety performance of construction projects: A literature review'. In: *IOP Conference Series: Materials Science and Engineering.* Vol. 713. 1. IOP Publishing, p. 012036.

Abdelbari, H. and Shafi, K. (2017). 'A computational intelligence-based method to 'learn'causal loop diagram-like structures from observed data'. In: *System Dynamics Review* 33.1, pp. 3–33.

Abioye, S. O. *et al.* (2021). 'Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges'. In: *Journal of Building Engineering* 44, p. 103299.

Aghaei, P., Asadollahfardi, G. and Katabi, A. (2021). 'Safety risk assessment in shopping center construction projects using Fuzzy Fault Tree Analysis method'. In: *Quality & Quantity*, pp. 1–17.

Ajayi, A. *et al.* (2020). 'Optimised big data analytics for health and safety hazards prediction in power infrastructure operations'. In: *Safety science* 125, p. 104656.

Albrechtsen, E. *et al.* (2018). *Forutseende sikkerhetsindikatorer i bygg-og anleggsbransjen.*

Alsheibani, S., Cheung, Y. and Messom, C. (2018). 'Artificial Intelligence Adoption: AI-readiness at Firm-Level.' In: *PACIS* 4, pp. 231–245.

Andreassen, E. *et al.* (2020). *Forutseende sikkerhetsindikatorer - Digitalisering i bygg og anlegg.* https://prosjektnorge.no/wp-content/uploads/2020/12/Digitalisering-Forutseende-sikkerhetsindikatorer.pdf.

Arbeidstilsynet (2023). *Forskjellen på HMS og SHA.* https://www.arbeidstilsynet.no/hms/hms-i-bygg-og-anlegg/forskjellen-pa-hms-og-sha/pageAsPdf.

Asilian-Mahabadi, H. *et al.* (2018). 'A qualitative investigation of factors influencing unsafe work behaviors on construction projects'. In: *Work* 61.2, pp. 281–293.

Baccarini, D. (1996). 'The concept of project complexity—a review'. In: *International journal of project management* 14.4, pp. 201–204.

Bala, B. K., Arshad, F. M. and Noh, K. M. (2017). 'System dynamics'. In: *Modelling and Simulation* 274.

Birisci, E., Gursakal, N. and Celik, S. (2023). *Synthetic Data for Deep Learning: Generate Synthetic Data for Decision Making and Applications with Python and R.* Berkeley, CA: Apress L. P.

Bonaccorso, G. (2017). *Machine learning algorithms.* Packt Publishing Ltd.

Borkenhagen, I. R. and Olsen, J. S. (2022). 'Machine Learning as a Leading Indicator for Predicting Safety Performance in the Construction Industry'.

Chen, T.-H. *et al.* (2022). 'Develop a Smart Microclimate Control System for Greenhouses through System Dynamics and Machine Learning Techniques'. In: *Water* 14.23, p. 3941.

Chen, Y.-T., Tu, Y.-M. and Jeng, B. (2011). 'A machine learning approach to policy optimization in system dynamics models'. In: *Systems Research and Behavioral Science* 28.4, pp. 369–390.

Choi, J. *et al.* (2020). 'Machine learning predictive model based on national data for fatal accidents of construction workers'. In: *Automation in Construction* 110, p. 102974.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches.* Sage publications.

Dangerfield, B. (2020). *System Dynamics.* Springer.

Direktoratet for forvaltning og økonomistyring (2022a). *Byggeprosessen.* https://anskaffelser.no/anskaffelsesprosessen/byggeprosessen/avklare-behov-i-bygg-og-anlegg.

— (2022b). *Utførelsesentreprise - BAE.* https://anskaffelser.no/hva-skal-du-kjope/bygg-anlegg-og-eiendom-bae/gjennomforingsmodeller/utforelsesentreprise-bae.

Duggan, J. (2020). 'Exploring the opportunity of using machine learning to support the system dynamics method: Comment on the paper by Edali and Yücel'. In: *Systems Research and Behavioral Science* 37.6, pp. 959–963.

Edali, M. and Yücel, G. (2020). 'Analysis of an individual-based influenza epidemic model using random forest metamodels and adaptive sequential sampling'. In: *Systems Research and Behavioral Science* 37.6, pp. 936–958.

Eikeland, P. (2001). 'Samspillet i byggeprosessen'. In: *Teoretisk analyse av byggeprosesser.*

Fellows, R. and Liu, A. (2003). *Research Methods for Construction.* Blackwell Science.

Feng, Y. and Trinh, M. T. (2019). 'Developing resilient safety culture for construction projects'. In: *Journal of construction engineering and management* 145.11, p. 04019069.

Forrester, J. W. (1994). 'System dynamics, systems thinking, and soft OR'. In: *System dynamics review* 10.2-3, pp. 245–256.

— (2007). 'System dynamics—the next fifty years'. In: *System Dynamics Review: The Journal of the System Dynamics Society* 23.2-3, pp. 359–370.

Gangadhari, R. K., Khanzode, V. and Murthy, S. (2022). 'Application of rough set theory and machine learning algorithms in predicting accident outcomes in the Indian petroleum industry'. In: *Concurrency and Computation: Practice and Experience* 34.26, e7277.

Gao, L., Lu, P. and Ren, Y. (2021). 'A deep learning approach for imbalanced crash data in predicting highway-rail grade crossings accidents'. In: *Reliability Engineering & System Safety* 216, p. 108019.

Gong, S. *et al.* (2021). 'Developing a dynamic supervision mechanism to improve construction safety investment supervision efficiency in China: Theoretical simulation of evolutionary game process'. In: *International Journal of Environmental Research and Public Health* 18.7, p. 3594.

Gregoriades, A. and Mouskos, K. C. (2013). 'Black spots identification through a Bayesian Networks quantification of accident risk index'. In: *Transportation Research Part C: Emerging Technologies* 28, pp. 28–43.

Han, S. *et al.* (2014). 'Toward an understanding of the impact of production pressure on safety performance in construction operations'. In: *Accident analysis & prevention* 68, pp. 106–116.

Hansen, G. (2019). *Samspillet i byggeprosessen.* Fagbokforlaget.

Hossin, M. and Sulaiman, M. N. (2015). 'A review on evaluation metrics for data classification evaluations'. In: *International journal of data mining & knowledge management process* 5.2, p. 1.

Huang, J. *et al.* (2022). 'An evolutionary game-theoretic analysis of construction workers' unsafe behavior: Considering incentive and risk loss'. In: *Frontiers in Public Health* 10.

Hussein, B. A. (2016). *Veien til suksess: fortellinger og refleksjoner fra reelle prosjektcaser*. Fagbokforlaget.

International Centre for Complex Project Management (2012). *Complex Project Manager Competency Standard*. Standard. International Centre for Complex Project Management.

ISO9000 (2015). *Quality management systems — Fundamentals and vocabulary*. Standard. Geneva, CH: International Organization for Standardization.

Iversen, L. C., Lilleland-Olsen, M. and Woldseth, M. K. (2016). 'TIDLIGFASE I BYGGEPROSJEKTER'. In: *BA2015*.

Jafari, P. *et al.* (2019). 'Leading safety indicators: Application of machine learning for safety performance measurement'. In: *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*. Vol. 36. IAARC Publications, pp. 501–506.

Japkowicz, N. and Shah, M. (2011). *Evaluating learning algorithms: a classification perspective*. Cambridge University Press.

Jiang, L. *et al.* (2021). 'Improvement of random forest by multiple imputation applied to tower crane accident prediction with missing data'. In: *Engineering, construction and architectural management*.

Jiang, X. *et al.* (2023). 'Using evolutionary game theory to study construction safety supervisory mechanism in China'. In: *Engineering, construction and architectural management* 30.2, pp. 514–537.

Jiang, Z., Fang, D. and Zhang, M. (2015). 'Understanding the causation of construction workers' unsafe behaviors based on system dynamics modeling'. In: *Journal of Management in Engineering* 31.6, p. 04014099.

Jin, Y. and Levitt, R. E. (1996). 'The virtual design team: A computational model of project organizations'. In: *Computational & Mathematical Organization Theory* 2, pp. 171–195.

Jing, W. *et al.* (2019). 'System dynamics modeling strategy for civil construction projects: The concept of successive legislation periods'. In: *Symmetry* 11.5, p. 677.

Jo, T. (2021). *Machine Learning Foundations*. Springer International Publishing.

Johansen, A. *et al.* (2019). *Project Risk and Opportunity Management: The Owner's Perspective*. Routledge.

Kang, K.-S., Koo, C. and Ryu, H.-G. (2022). 'An interpretable machine learning approach for evaluating the feature importance affecting lost workdays at construction sites'. In: *Journal of Building Engineering* 53, p. 104534.

Kang, K. and Ryu, H. (2019). 'Predicting types of occupational accidents at construction sites in Korea using random forest model'. In: *Safety Science* 120, pp. 226–236.

Khalid, U., Sagoo, A. and Benachir, M. (2021). 'Safety Management System (SMS) framework development–Mitigating the critical safety factors affecting Health and Safety performance in construction projects'. In: *Safety science* 143, p. 105402.

Kim, G. and Lim, S. (2022). 'Development of an Interpretable Maritime Accident Prediction System Using Machine Learning Techniques'. In: *IEEE Access* 10, pp. 41313–41329.

Kim, J.-M. *et al.* (2022). 'A Deep Learning Model Development to Predict Safety Accidents for Sustainable Construction: A Case Study of Fall Accidents in South Korea'. In: *Sustainability* 14.3, p. 1583.

Kjellen, U. and Albrechtsen, E. (2017). *Prevention of accidents and unwanted occurrences: Theory, methods, and tools in safety management.* CRC Press.

Koc, K., Ekmekcioğlu, Ö. and Gurgun, A. P. (2021). 'Integrating feature engineering, genetic algorithm and tree-based machine learning methods to predict the post-accident disability status of construction workers'. In: *Automation in Construction* 131, p. 103896.

— (2022). 'Accident prediction in construction using hybrid wavelet-machine learning'. In: *Automation in Construction* 133, p. 103987.

Kvålshaugen, R. and Swärd, A. (2018). *Sluttrapport "Samhandling og prosessledelse"*. Tech. rep. Handelshøyskolen BI.

Lee, J. Y. *et al.* (2020). 'A study on data pre-processing and accident prediction modelling for occupational accident analysis in the construction industry'. In: *Applied Sciences* 10.21, p. 7949.

Lereim, J. (2013). *Kvalitetsledelse i prosjekter: en erfaringsbasert tilnærming.* Akademika forlag.

Li, J., Alburaikan, A. and Fátima Muniz, R. de (2023). 'Evaluation of safety-based performance in construction projects with neutrosophic data envelopment analysis'. In: *Management Decision* 61.2, pp. 552–568.

Li, X. *et al.* (2022). 'Understanding the influence of safety climate and productivity pressure on non-helmet use behavior at construction sites: A case study'. In: *Engineering, Construction and Architectural Management* 29.1, pp. 72–90.

Lowhorn, G. L. (2007). 'Qualitative and quantitative research: How to choose the best design'. In: *Academic Business World International Conference.*

Man, S. S. *et al.* (2021). 'The effect of personal and organizational factors on the risk-taking behavior of Hong Kong construction workers'. In: *Safety science* 136, p. 105155.

Maruyama, M. (1963). 'The second cybernetics: Deviation-amplifying mutual causal processes'. In: *American scientist* 51.2, pp. 164–179.

McCulloch, W. S. and Pitts, W. (1943). 'A logical calculus of the ideas immanent in nervous activity'. In: *The bulletin of mathematical biophysics* 5, pp. 115–133.

Mohamed, S. and Chinda, T. (2011). 'System dynamics modelling of construction safety culture'. In: *Engineering, Construction and Architectural Management.*

Mohammadi, A. and Tavakolan, M. (2019). 'Modeling the effects of production pressure on safety performance in construction projects using system dynamics'. In: *Journal of safety research* 71, pp. 273–284.

Mohammadi, A., Tavakolan, M. and Khosravi, Y. (2018). 'Factors influencing safety performance on construction projects: A review'. In: *Safety science* 109, pp. 382–397.

Mohammed, M., Khan, M. B. and Bashier, E. B. M. (2016). *Machine learning: algorithms and applications.* Crc Press.

Mostue, B. *et al.* (2022). *Ulykker i bygg og anlegg - Rapport 2022.* https://www.arbeidstilsynet.no/contentassets/1715bdd4ec5943358b024e206969a5d4/kompass-rapport-01-2022-ulykker-bygg-og-anlegg.

Muñoz-La Rivera, F., Mora-Serrano, J. and Oñate, E. (2021). 'Factors influencing safety on construction projects (FSCPs): types and categories'. In: *International journal of environmental research and public health* 18.20, p. 10884.

Nayak, N. R. *et al.* (2022). 'Network mining techniques to analyze the risk of the occupational accident via bayesian network'. In: *International Journal of System Assurance Engineering and Management* 13.1, pp. 633–641.

Newman, I., Benz, C. R. and Ridenour, C. S. (1998). *Qualitative-quantitative research methodology: Exploring the interactive continuum.* SIU Press.

Nikolenko, S. I. (2021). *Synthetic data for deep learning.* Vol. 174. Springer.

NTNU (n.d.). *Sustainable value creation by digital predictions of safety performance in the construction industry - DiSCo.* https://www.ntnu.edu/iot/sustainable-value-creation-by-digital-predictions-of-safety-performance-in-the-construction-industry-disco-.

Olsson, N. O. and Bull-Berg, H. (2015). 'Use of big data in project evaluations'. In: *International Journal of Managing Projects in Business.*

Olsson, N. O., Hansen, G. K. and Blakstad, S. H. (2022). 'Who are the users? User categorisations and implications on building performance measurement'. In: *International Journal of Public Sector Performance Management* 10.4, pp. 566–579.

PMI Standards Committee and Project Management Institute (1996). 'A guide to the project management body of knowledge'. In: Project Management Institute.

Poh, C. Q., Ubeynarayana, C. U. and Goh, Y. M. (2018). 'Safety leading indicators for construction sites: A machine learning approach'. In: *Automation in construction* 93, pp. 375–386.

Prabhakaran, V. and Martin Jr, D. (2020). 'Participatory machine learning using community-based system dynamics'. In: *Health and Human Rights* 22.2, p. 71.

Project Management Institute, Inc. (2016). *Construction extension to the PMBOK guide.* Project Management Institute, Inc.

— (2017). *A guide to the Project Management Body of Knowledge (PMBOK® Guide).* Project Management Institute, Inc.

Qayoom, A. and HW Hadikusumo, B. (2019). 'Multilevel safety culture affecting organization safety performance: a system dynamic approach'. In: *Engineering, Construction and Architectural Management* 26.10, pp. 2326–2346.

Ramachandran, K. *et al.* (2022). 'Machine learning and role of artificial intelligence in optimizing work performance and employee behavior'. In: *Materials Today: Proceedings* 51, pp. 2327–2331.

Rausand, M. (2013). *Risk assessment: theory, methods, and applications.* Vol. 115. John Wiley & Sons.

Recal, F. and Demirel, T. (2021). 'Comparison of machine learning methods in predicting binary and multi-class occupational accident severity'. In: *Journal of Intelligent & Fuzzy Systems* 40.6, pp. 10981–10998.

Regona, M. *et al.* (2022). 'Opportunities and adoption challenges of AI in the construction industry: a PRISMA review'. In: *Journal of open innovation: technology, market, and complexity* 8.1, p. 45.

Rodrigues, A. G. and Williams, T. M. (1997). 'System dynamics in software project management: towards the development of a formal integrated framework'. In: *European Journal of Information Systems* 6.1, pp. 51–66.

— (1998). 'System dynamics in project management: assessing the impacts of client behaviour on project performance'. In: *Journal of the operational research society* 49.1, pp. 2–15.

Rolstadås, A. *et al.* (2020). *Praktisk prosjektledelse: fra idé til gevinst.* Second edition. Fagbokforlaget.

Roozkhosh, P., Pooya, A. and Agarwal, R. (2022). 'Blockchain acceptance rate prediction in the resilient supply chain with hybrid system dynamics and machine learning approach'. In: *Operations Management Research*, pp. 1–21.

Rostamnezhad, M. *et al.* (2020). 'Modeling social sustainability in construction projects by integrating system dynamics and fuzzy-DEMATEL method: a case study of highway project'. In: *Engineering, construction and architectural management* 27.7, pp. 1595–1618.

Russell, S. and Norvig, P. (2016). *Artificial Intelligence: A Modern Approach.* Pearson Education, Inc.

Sadeghi, H. *et al.* (2020). 'Developing an ensemble predictive safety risk assessment model: Case of Malaysian construction projects'. In: *International journal of environmental research and public health* 17.22, p. 8395.

Sarkar, S. *et al.* (2019). 'Application of optimized machine learning techniques for prediction of occupational accidents'. In: *Computers & Operations Research* 106, pp. 210–224.

Sarkar, S. *et al.* (2020). 'Predicting and analyzing injury severity: A machine learning-based approach using class-imbalanced proactive and reactive data'. In: *Safety science* 125, p. 104616.

Shin, Y. (2019). 'Application of stochastic gradient boosting approach to early prediction of safety accidents at construction site'. In: *Advances in Civil Engineering* 2019.

Smith, C. *et al.* (2006). *The history of artificial intelligence.* https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf.

Statistisk Sentralbyrå (2022). *Arbeidsulykker.* https://www.ssb.no/statbank/table/10913/tableViewLayout1/.

Sterman, J. (2013). *Introduction To System Dynamics.* https://www.youtube.com/watch?v=AnTwZVviXyY. MIT.

Stock, J. H. and Watson, M. W. (2008). *Business cycles, indicators, and forecasting.* University of Chicago Press.

Su, W. *et al.* (2021). 'Developing a construction safety standard system to enhance safety supervision efficiency in China: A theoretical simulation of the evolutionary game process'. In: *Sustainability* 13.23, p. 13364.

Sukamani, D. and Wang, J. (2020). 'SEM Model for Investigating Factor of an Accident Affecting Safety Performance in Construction Sites in Nepal.' In: *Engineering Letters* 28.3.

Sultana, S., Andersen, B. S. and Haugen, S. (2019). 'Identifying safety indicators for safety performance measurement using a system engineering approach'. In: *Process Safety and Environmental Protection* 128, pp. 107–120.

Sun, X. *et al.* (2019). 'A system dynamics model of prevention through design towards eliminating human error'. In: *KSCE Journal of Civil Engineering* 23, pp. 1923–1938.

Sweetser, A. (1999). 'A comparison of system dynamics (SD) and discrete event simulation (DES)'. In: *17th International Conference of the System Dynamics Society*, pp. 20–23.

Tiltnes, S. (2015). *Veileder for Fasenormen "Neste Steg" - et Felles Rammeverk for Norske Byggeprosesser*. https://bygg21.no/wp-content/uploads/2021/03/veileder-for-stegstandard-ver-1.2-med-logoer-201116.pdf.

Turing, A. M. (2009). *Computing machinery and intelligence*. Springer.

University of Bergen (2022). *What is system dynamics?* https://www.uib.no/en/rg/dynamics/39282/what-system-dynamics#:~:text=System%20Dynamics%20is%20a%20computer,the%20other%20Systems%20Thinking%20approaches.

Usukhbayar, R. and Choi, J. (2020). 'Critical safety factors influencing on the safety performance of construction projects in Mongolia'. In: *Journal of Asian Architecture and Building Engineering* 19.6, pp. 600–612.

Wade, C. and Glynn, K. (2020). *Hands-On Gradient Boosting with XGBoost and scikit-learn: Perform accessible machine learning and extreme gradient boosting with Python*. Packt Publishing Ltd.

Wang, Q. (2019). 'How to apply AI technology in Project Management'. In: *PM World Journal*.

Weng, T. *et al.* (2022). 'Modeling chaotic systems: Dynamical equations vs machine learning approach'. In: *Communications in Nonlinear Science and Numerical Simulation* 114, p. 106452.

Winch, G. M. (2010). *Managing construction projects*. Second edition. John Wiley & Sons.

Xie, H. *et al.* (2019). 'Effects of payment delays at two links in payment chains on the progress of construction projects: system dynamic modeling and simulation'. In: *Sustainability* 11.15, p. 4115.

Yap, J. B. H. and Lee, W. K. (2020). 'Analysing the underlying factors affecting safety performance in building construction'. In: *Production Planning & Control* 31.13, pp. 1061–1076.

Yedla, A., Kakhki, F. D. and Jannesari, A. (2020). 'Predictive modeling for occupational safety outcomes and days away from work analysis in mining operations'. In: *International journal of environmental research and public health* 17.19, p. 7054.

Zhang, S. *et al.* (2022). 'Hybrid feature selection-based machine learning Classification system for the prediction of injury severity in single and multiple-vehicle accidents'. In: *PLoS one* 17.2, e0262941.

Zhou, Z.-H. (2021). *Machine learning*. Springer Nature.

Zhu, Z. *et al.* (2020). 'Developing key safety management factors for construction projects in China: A resilience perspective'. In: *International Journal of Environmental Research and Public Health* 17.17, p. 6167.

Appendix

A Correlation Matrices

Serious Accident Rate Dataset

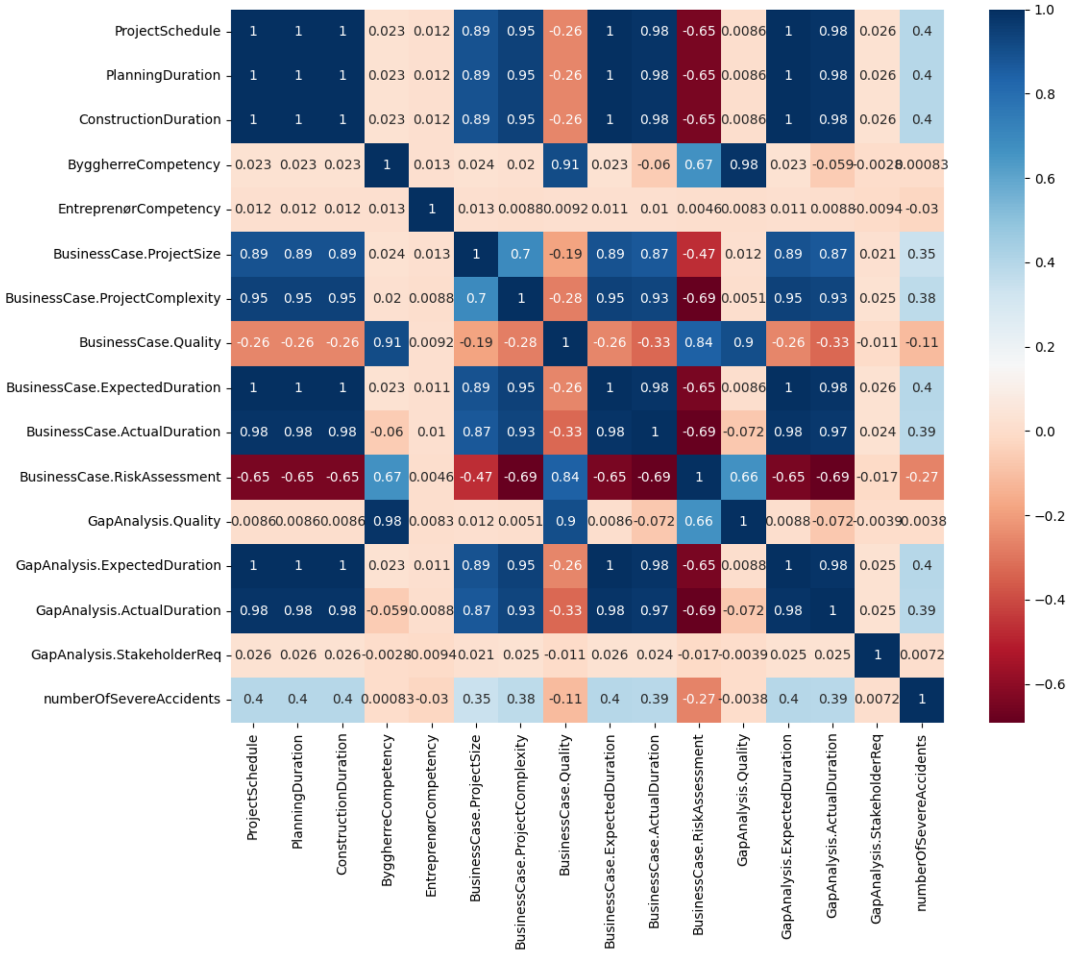


Figure A.1: Serious accidents correlation matrix for strategic definition stage

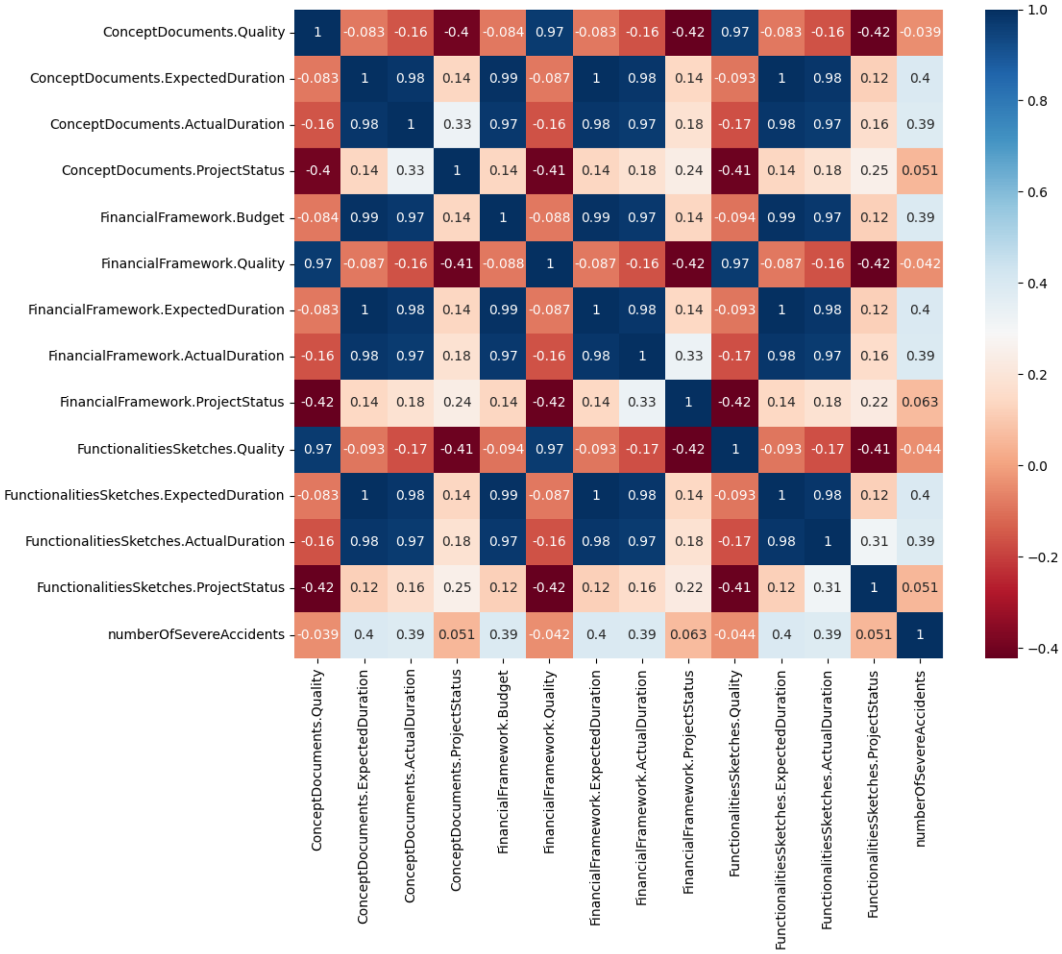


Figure A.2: Serious accidents correlation matrix for concept development stage

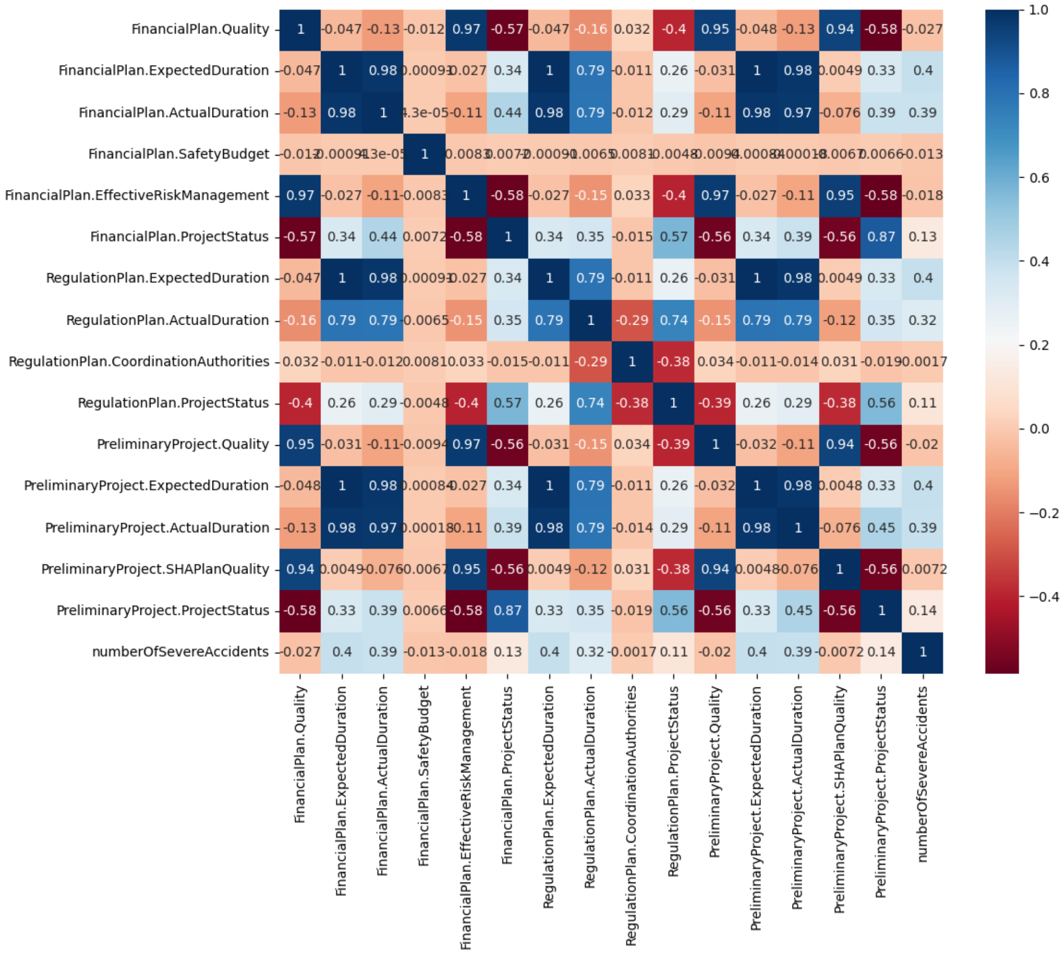


Figure A.3: Serious accidents correlation matrix for concept processing stage

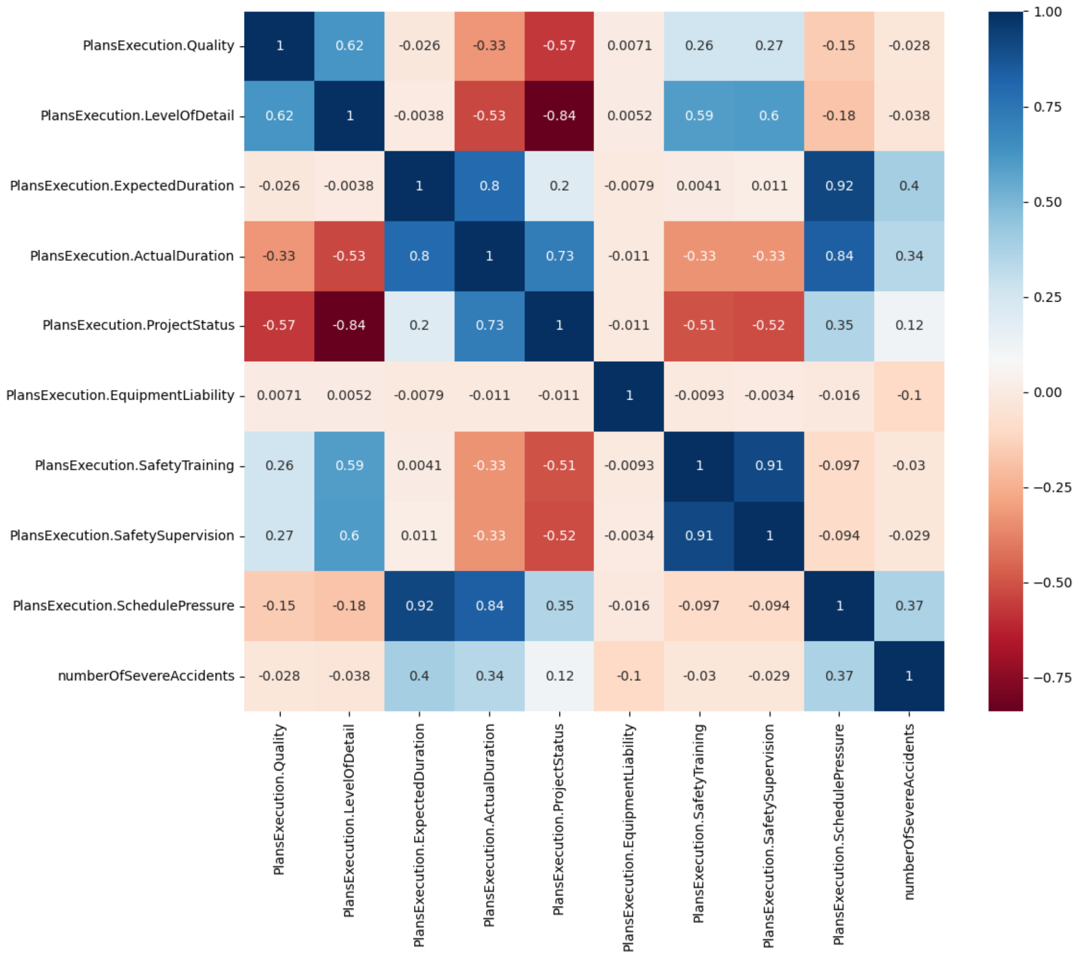


Figure A.4: Serious accidents correlation matrix for detailed engineering stage

Fatal Accident Rate Dataset

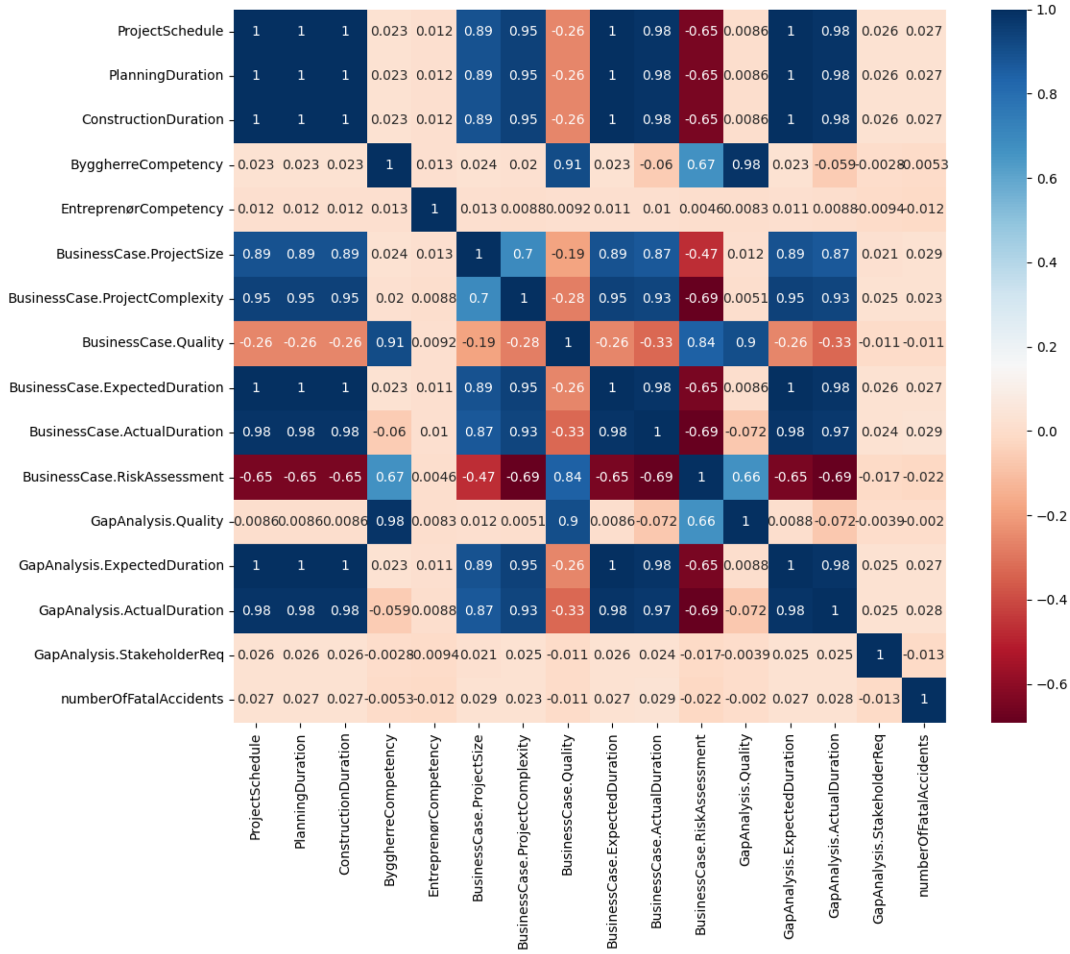


Figure A.5: Fatal accidents correlation matrix for strategic definition stage

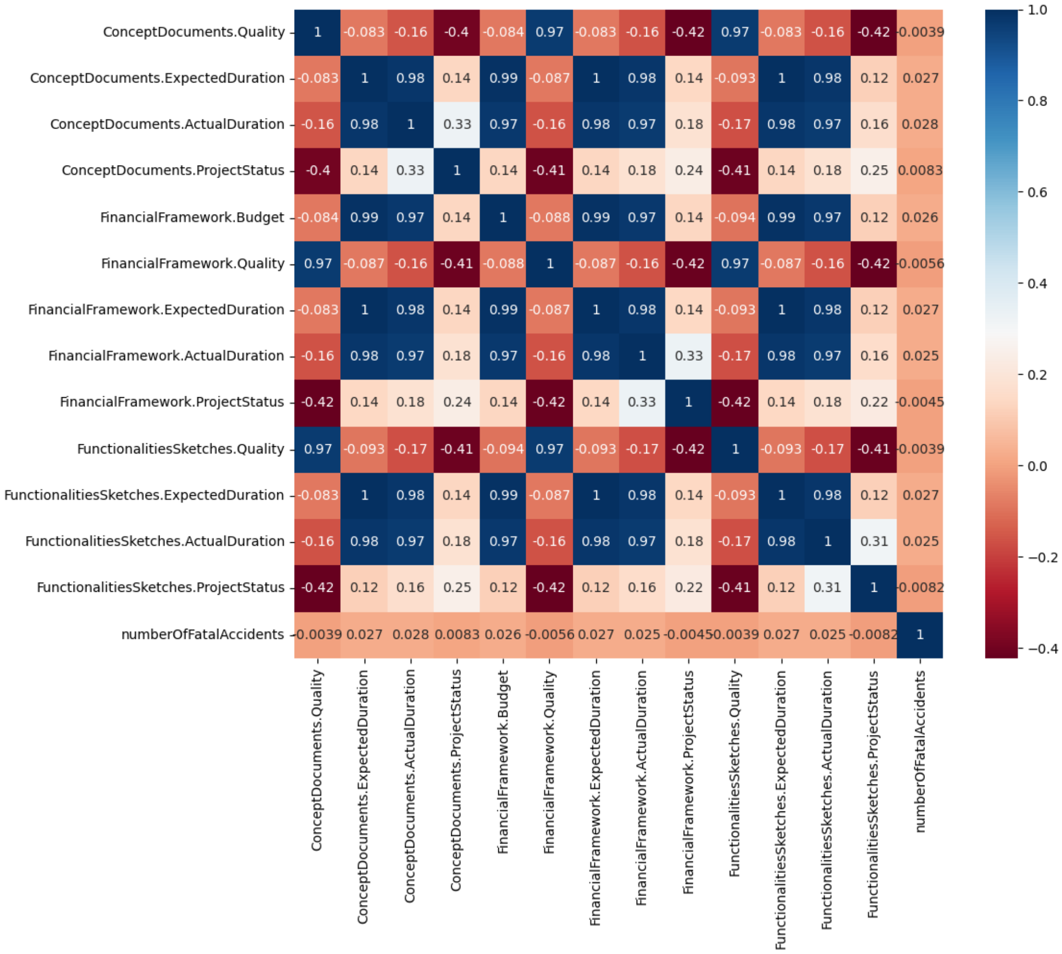


Figure A.6: Fatal accidents correlation matrix for concept development stage

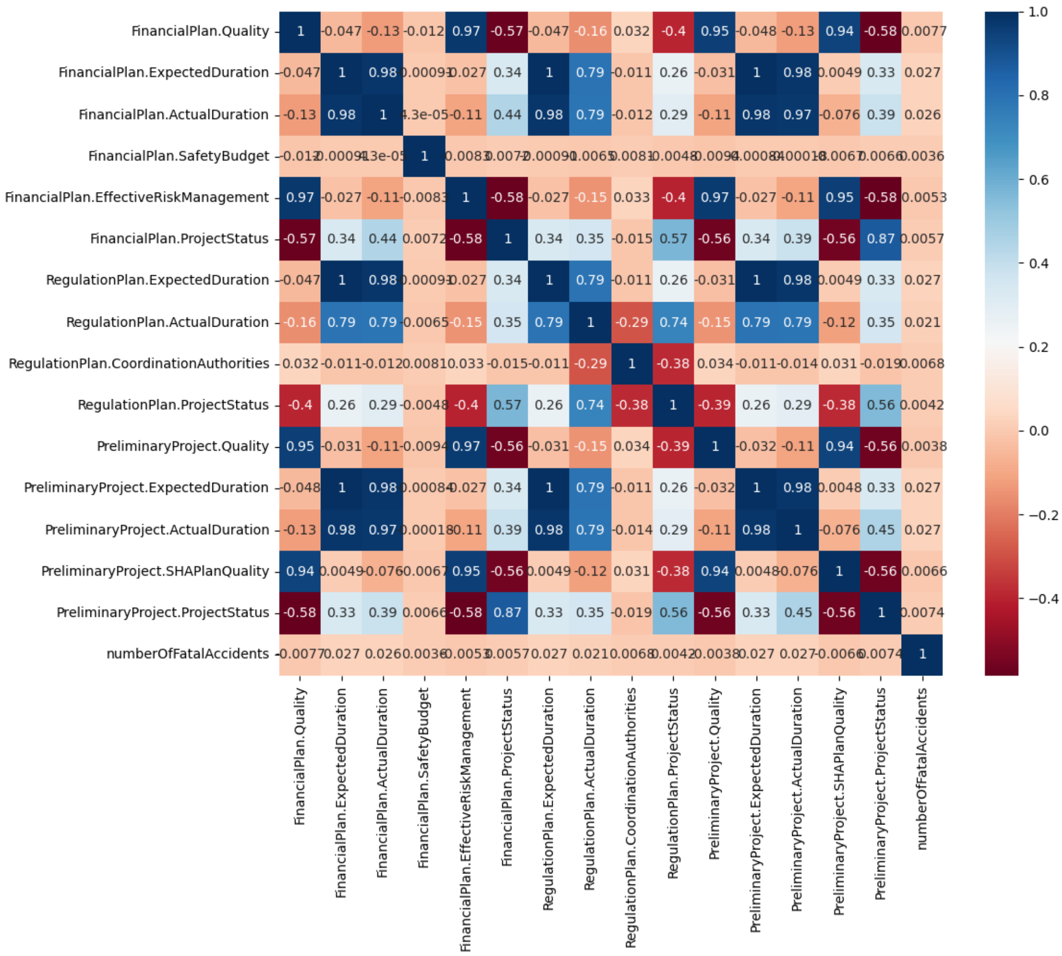


Figure A.7: Fatal accidents correlation matrix for concept processing stage

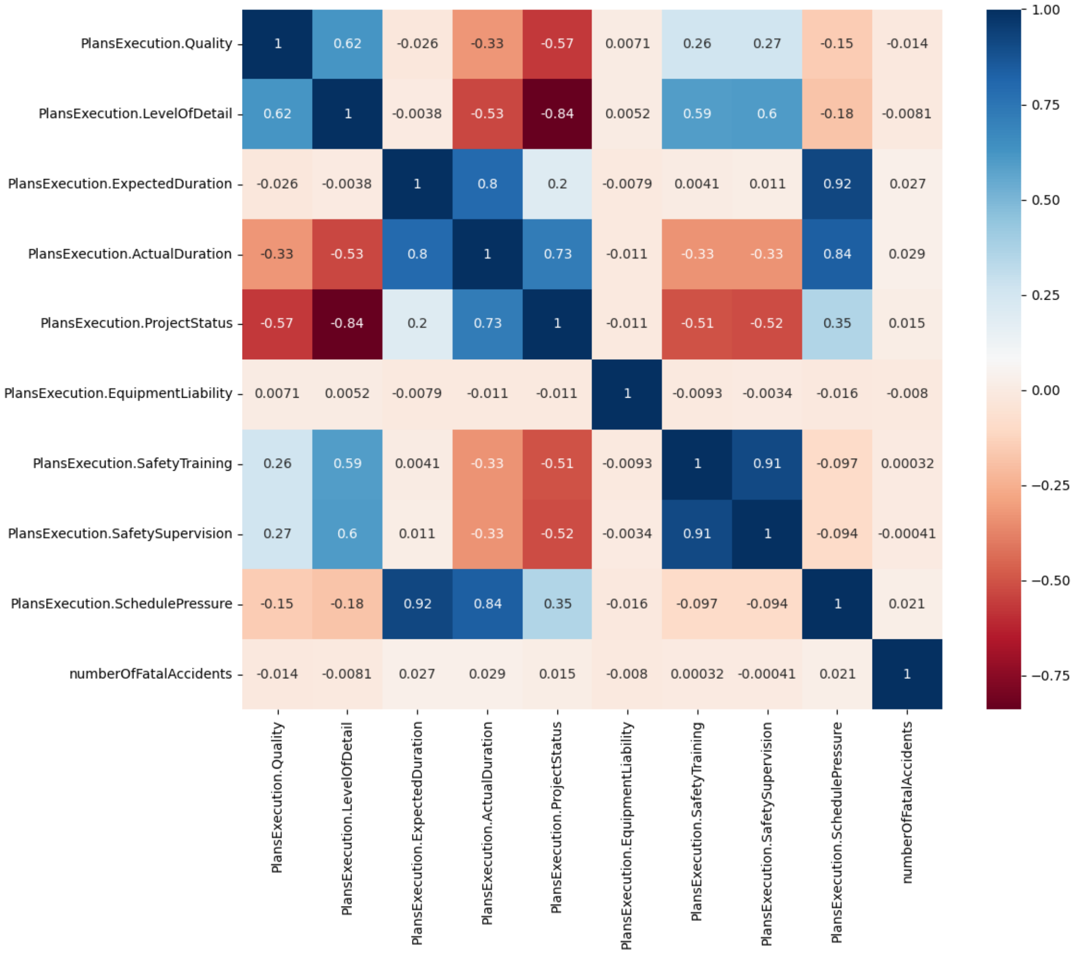


Figure A.8: Fatal accidents correlation matrix for detailed engineering stage

NTNU

Norwegian University of
Science and Technology