Master's thesis

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

Maud Johansson
Anne Høgdahl Skjærseth

Utilizing machine learning for lead time prediction in the MTO fenestration industry

Master's thesis in Engineering and ICT
Supervisor: Erlend Alfnes
Co-supervisor: Maria Flavia Mogos and Maryna Waszak
June 2023

NTNU
Norwegian University of Science and Technology

Maud Johansson
Anne Høgdahl Skjærseth

Utilizing machine learning for lead time prediction in the MTO fenestration industry

Master's thesis in Engineering and ICT
Supervisor: Erlend Alfnes
Co-supervisor: Maria Flavia Mogos and Maryna Waszak
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 master's thesis is submitted as the final work of the Master of Engineering and ICT at the department of Production Management at the Norwegian University of Science and Technology.

We would like to thank our supervisor Erlend Alfnes, Associate Professor at the Department of Mechanical and Industrial Engineering at NTNU, for his guidance and feedback during this year, which has been a great contribution to this work.

We would like to thank our co-supervisor's from Sintef Manufacturing, Maria Flavia Mogos, and Sintef Digital, Maryna Waszak. They have both provided us with great knowledge and insight in their research fields, industrial ecosystems, and sustainable communication technologies. Additionally, they have shared important insight about the case company of this project, which has been in corporation with Sintef for several years.

We would like to thank Gilje, the case company, which has been a great cooperation partner through this thesis. Special thanks to Karl Inge Lygre, Kjetil Gilje, Tor Steinar Gilje and the rest of our contact people at Gilje, which has taken the time to let us visit their factory, have meetings and answer other questions that we have had throughout the project. In addition, we would like to thank Nordic Door and MagnorVidnuet for their time and shared knowledge in the interviews we had with their representatives.

We would like to thank our friends, boyfriends and families for support and motivation, and specially to Anne's father who has provided great feedback to the thesis. Last, but definitely not least, we would like to thank our fellow students for great discussions and joyful conversations. Specially our two classmates which we have spent the whole year in the same room with at PFI.

Abstract

The fenestration industry experience increasing demand for customized products, shifting market trends, and stricter regulations for energy-efficient windows. In order to meet these requirements the companies are dependent on realistic and flexible production plans and production control. A key parameter to achieve this is the lead time. Most fenestration companies operate in a make to order (MTO) environment, which is a planning strategy that begins production after a customer order is placed. The demand uncertainty and high product variety in MTO fenestration industry makes the task of estimating lead times difficult, and increases the necessity of lead time prediction. The utilization of machine learning (ML) for lead time prediction has received increased attention in the literature. In addition, increasing regularity of implementing technology for data collection, such as Radio-Frequency Identification (RFID) systems, in manufactures enhances the availability of utilizing ML for lead time prediction.

This thesis aims to *contribute to increased understanding of how ML-based lead time prediction can be developed and integrated to support production planning and control (PPC) in the MTO fenestration industry.* An industry independent process model called CRISP-DM was applied to achieve this goal. This process consists of six steps; business understanding, data understanding, data preparation, data modelling, data evaluation and deployment.

To achieve the thesis goal three research questions (RQs) were presented: 1) In what areas of production planning and control can lead time prediction be applied in MTO fenestration companies? 2) What can be the causes for data quality issues in data applied for lead time prediction? 3) How can CRISP-DM be applied to predict lead times in the MTO fenestration industry?

The methods applied to answer the RQs included literature reviews, interviews with three different fenestration companies, case study, data analysis and data modelling. RQ1 was answered in terms of literature, case company and multi-company interviews. RQ2 was addressed through the literature, as well as data analysis and insight from the case company. RQ3 was achieved through data modelling on the collected data from the case company, and by combining the answers from RQ1 and RQ2.

Five application areas for lead time prediction were identified in the literature. These include *capacity planning*, *due date setting*, *production scheduling*, *production analysis*, and *production monitoring*. The multi-company interviews revealed varying levels of interest in adopting lead time prediction for the application areas. The two larger companies highlighted capacity planning, production analysis and production monitoring as the most relevant areas for LTP within PPC. The smaller company showed low interest across all areas identified in the literature. The size and digitization level of the companies were highlighted as possible causes for the dispersed answers.

A data quality assessment was performed in order to detect data quality issues in the collected

data. The assessment's primary finding was 40% inconsistent RFID shop floor scannings, and contradictory values between data sources. Causes were detected from the literature and from the case company, which included several errors regarding the RFID scanners and tags, in addition to operators ignorance and negligence. The detected errors were later removed during data preparation, in order for the data to be utilized as the input dataset for the ML models.

A process model was developed in order to address all six phases of the CRISP-DM model for lead time prediction in MTO fenestration companies. The case study and multi-company interviews gave the foundation for *business understanding*. *Data understanding* was achieved through the data quality assessment from RQ2, where the detected errors were removed during *data preparation*. Two ML models were developed for *modelling*, namely Random forest (RF) and a multi-layer perceptron (MLP) model. The *evaluation* showed that the RF model outperformed the MLP model in all performance metrics. Lastly, the *deployment* was not put in practice, but the potential application areas in PPC, detected in RQ1, were addressed and examined.

The thesis main contribution to theory is an extension of the existing CRISP-DM framework by developing an industry specific process model for ML-based lead time prediction in the MTO fenestration industry. The contribution to practice is a guide for which develops a shared understanding between data scientists and production managers. This can assist companies in effectively integrating lead time prediction into their PPC processes, thereby enhancing operational planning and decision-making. The main limitation to the study is that the data foundation for the process model was only collected from one MTO fenestration company. Future research should therefore apply the process model to other fenestration companies to test it's robustness and generalizability.

Sammendrag

Vindu- og dørindustrien opplever økende etterspørsel etter kundetilpassede produkter, skiftende markedstrender og strengere forskrifter for energieffektive vinduer. For å imøtekomme disse kravene er selskapene avhengige av realistiske og fleksible produksjonsplaner og produksjonskontroll. En viktig parameter for å oppnå dette er ledetiden. De fleste vindu- og dørbedrifter opererer i et kundetilpasset miljø, "make to order" (MTO), som er en planleggingsstrategi som starter produksjonen etter at en kundeordre er plassert. Uforutsigbar etterspørsel og høy produktvariasjon i vindu- og dørindustrien gjør det vanskelig å estimere ledetider, og øker behovet for ledetidsprediksjon. Bruken av maskinlæring (ML) for ledetidsprediksjon har fått økt oppmerksomhet i litteraturen. I tillegg blir teknologier for datainnsamling, som Radio-Frequency Identification (RFID) systemer, mer utbredt i produksjonsprosesser, noe som øker muligheten for å anvende maskinlæring til ledetidspredikering.

Denne masteroppgaven har som mål å *bidra til økt forståelse for hvordan ML-basert ledetidsprediksjon kan utvikles og integreres for å støtte produksjonsplanlegging og kontroll (PPC) i den kundetilpassede vindu- og dørindustrien.* En industriuavhengig prosessmodell kalt CRISP-DM ble brukt for å oppnå dette målet. Denne prosessen består av seks trinn: bedriftsforståelse, dataforståelse, dataprosessering, datamodellering, dataevaluering og integrering.

For å oppnå oppgavens mål ble det presentert tre forskningsspørsmål (RQ): 1) I hvilke områder av produksjonsplanlegging og -kontroll kan ledetidsprediksjon brukes i kundetilpassede vindu- og dørbedrifter? 2) Hva kan være årsakene til datakvalitetsproblemer i anvendt data for ledetidsprediksjon? 3) Hvordan kan CRISP-DM brukes til å predikere ledetider i den kundetilpassede vindu- og dørindustrien?

Metodene som ble brukt for å besvare RQene inkluderte litteraturstudier, intervjuer med tre forskjellige vindu- og dørbedrifter, case-studie, dataanalyse og datamodellering. RQ1 ble besvart ved bruk av litteraturstudie, case-selskapet og bedriftsintervjuene. RQ2 ble adressert gjennom litteraturen, samt dataanalyse og innsikt fra case-selskapet. RQ3 ble oppnådd gjennom datamodellering basert på innsamlet data fra case-selskapet, og ved å kombinere svarene fra RQ1 og RQ2.

Fem bruksområder for ledetidsprediksjon ble identifisert i litteraturen. Disse inkluderer *kapasitetsplanlegging*, *fastsettelse av leveringsdato*, *operasjonsplanlegging*, *produksjonsanalyse* og *produksjonsovervåking*. Bedriftsintervjuene avslørte varierende interesse for å implementere ledetidsprediksjon for disse bruksområdene. De to større selskapene fremhevet kapasitetsplanlegging, produksjonsanalyse og produksjonsovervåking som de mest relevante områdene for ledetidsprediksjon innenfor PPC. Det mindre selskapet viste lav interesse for alle områdene som ble identifisert i litteraturen. Størrelsen og digitaliseringsnivået til selskapene ble fremhevet som mulige årsaker til de forskjellige svarene.

En vurdering av datakvalitet ble utført for å undersøke problemer i den innsamlede dataen. Vurderingen avdekket 40% inkonsistente RFID skanninger på fabrikkgulvet og motstridende verdier mellom datakildene. Årsakene ble identifisert gjennom litteraturen og fra case-selskapet, som inkluderte flere feil relatert til RFID-skannere og tagger, i tillegg til operatørenes uvitenhet og uaktsomhet. De oppdagede feilene ble senere fjernet under dataprosesseringen, slik at dataen kunne brukes som input til ML-modellene.

En prosessmodell som inkluderte alle seks faser av CRISP-DM-modellen ble utviklet for ledetidsprediksjon i kundetilpassede vindu- og dørbedrifter. Case-studiet og bedriftsintervjuene dannet grunnlaget for *bedriftsforståelse*. *Dataforståelsen* ble oppnådd gjennom vurderingen av datakvalitet fra RQ2, der de oppdagede feilene ble fjernet under *dataprosesseringen*. To ML-modeller ble utviklet for *modellering*, nemlig Random Forest (RF) og en multi-layer perceptron (MLP)-modell. *Evalueringen* viste at RF-modellen presterte bedre enn MLP-modellen i alle ytelsesmålene. *Integrering* av resultatene ble ikke satt ut i praksis, men de potensielle bruksområdene i PPC som ble identifisert i RQ1, ble adressert og undersøkt.

Det teoretiske hovedbidraget til masteroppgaven er en utvidelse av det eksisterende CRISP-DM-rammeverket ved å utvikle en bransjespesifikk prosessmodell for ML-basert ledetidsprediksjon i den kundetilpassede vindu- og dørindustrien. Bidraget til praksis er retningslinjer som kan skape en felles forståelse mellom dataforskere og produksjonsledere. Dette kan hjelpe selskaper med å integrere ledetidsprediksjon på en effektiv måte i deres PPC-prosesser, og dermed forbedre operativ planlegging og beslutningstaking. Den største begrensningen i studiet er at datagrunnlaget for prosessmodellen bare ble samlet inn fra ett vinudsselskap. Fremtidig forskning bør derfor anvende prosessmodellen på andre vindu- og dørbedrifter for å teste dens robusthet og generaliserbarhet.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

R^2 Coefficient of determination. 73

Acc Accuracy. 81

AI Artificial intelligence. 30

ANN Artificial neural network. 33

ATO Assemble to order. 19

AutoML Automated machine learning. 35

BD Big data. 41

BDT Boosting decision tree. 33

Comp Completeness. 81

Cons Consistency. 81

CPS Cyber-physical system. 45

CRISP-DM The Cross-industry standard process for data mining. 3

CRP Capacity requirement planning. 20

DNN Deep neural network. 33

DT Decision tree. 33

ETO Engineer to order. 6

IoT Internet of things. 45

IQR Interquartile range. 64

k-NN k-nearest neighbor. 33

LassoReg Lasso regression. 33

LR Linear regression. 33

LTP Lead time prediction. 3

MAE Mean absolute error. 73

MARS Multivariate adaptive regression. 33

ML Machine learning. 3

MLP Multi-layer perceptron. 39

MLR Multinomial logistic regression. 34

MPS Master production schedule. 20

MRP Material requirements planning. 20

MSE Mean squared error. 73

MTO Make to order. 1

MTS Make to stock. 1

NBC Naive bayes classifier. 34

ODCP Order decoupling point. 19

PC Principle component. 70

PCA Principle component analysis. 70

PCC Pearson correlation coefficient. 68

PPC Production planning and control. 1

RF Random forest. 33

RFID Radio-frequency identification. 3

RidgeReg Ridge regression. 33

RMSE root mean squared error. 73

RQ Research question. 5

SR Simple ratio. 44

SVR Support vector regression. 33

WIP Work-in-process. 35

WLC Workload control. 34

1 Introduction

1.1 Problem background

Window manufacturers claim that windows are becoming less of a utility element in a home and more a design object, which increases the demand of customized products. Higher energy requirements for windows also put a greater focus on innovative ideas and new manufacturing requirements. The fenestration industry is a collective term for manufactures producing windows, doors, and glass. This and other industries are being pushed to adopt a make to order (MTO) approach due to the rising demand of customized products (C. Wang and Jiang 2019). MTO is defined by a planning strategy that begins production after a customer order is received, in contrast to make to stock (MTS) industries such as smartphone production, which produces based on forecast. An MTO planning strategy begins with the order fulfillment process, which includes steps for the design of the product, the production process, delivery, and reporting of order progress (Cox and Blackstone 2002). Uncertainty in the demand in MTO environments makes production planning and control (PPC), which involves the crucial tasks of organizing, planning and performing manufacturing activities, a complex and difficult task (Cañas, Mula et al. 2022).

The building industry has an important role in order to achieve the national and global objective of creating a sustainable future. According to Council (2019), buildings account for 39% of all energy-related carbon emissions. The energy consumption of a building is significantly influenced by the building envelope (Misiopecki, Bouquin et al. 2018). This results in stricter national and international regulations, such as EU Energy Performance in Buildings Directive (EU 2023). In Norway, the building sector has seen consistent expansion over the past few decades. Building construction climbed by 5.9% and construction project development by 4.2% in 2022, respectively, according to Statics Norway (SSB 2023). The need of manufacturing high-performance, energy-efficient windows is growing, necessitating a swift restructuring of the fenestration sector (Blackston 2022). Manufacturers must adhere to international regulations in addition to taking market trends and client demands into account. Today's market trend for windows is larger glass panes, frames, and casement sizes (Misiopecki, Bouquin et al. 2018). The buyer also wants a variety of designs, including different colors and shapes of the frames, as well as casements at affordable costs. Gilje Tre AS, a Norwegian window company, claims that increased competition from East European nations pushes the manufactures to produce high-quality windows more quickly and at a lower cost, including energy cost. This competition has been identified in several other Norwegian sectors as well (Regjeringen 2001).

Window companies must conform to the frequent changes in their product range in order to meet the growing demand for customized goods, energy-efficient windows, and shifting market trends. These characteristics affect several aspects of manufacturing. Production flexibility is required in order to adjust the production processes and accommodate different specifications and variations.

Increased manufacturing efficiency is required in order to adopt to the advanced production processes seen in MTO companies. Additionally, overall efficient supply chain management is required to ensure timely availability of resources and material for the continuous stream of incoming orders. MTO companies are dependent on realistic and flexible production plans to easily track the progress of the order to ensure reliable delivery dates to the customers (Kapulin and Russkikh 2020). Products within the MTO fenestration industry form an important part of buildings and are thus an important element within the construction industry. Nevertheless, the authors found production management within the fenestration industry to be a research field which has received little attention in previous literature.

Lead time is an important performance indicator for MTO companies, through exploring and reducing the lead time companies' can gain competitive advantage in today's dynamically changing market (Gyulai, Pfeiffer, Bergmann et al. 2018; Lim, Yusof et al. 2019; Mikati 2010; Mourtzis, Doukas et al. 2014). Lead time is the time elapsed from the release of an order to the production floor and its placement in the finished goods inventory (Szaller, Béres et al. 2018). The lead time can indicate the overall performance of the company, and it has crucial impact on the company's customer relationship. It also reveals the adaptability and productivity of the manufacturer. Lead times of incoming orders can be utilized in important PPC tasks in order to develop flexible and realistic operational plans. Reduction in the lead time can increase the amount of sales, lower the production costs, and save energy which contributes to a more sustainable production (Asadzadeh, Azadeh et al. 2011; Mikati 2010). From the viewpoint of the consumer, the lead time affects when they receive their product, namely the delivery date, and on-time delivery is a crucial parameter for customer satisfaction (Mourtzis, Doukas et al. 2014; Öztürk, Kayalıgil et al. 2006). In the fenestration sector, orders are frequently placed by customers weeks or months before they want them delivered, and the manufacturer must ensure prompt delivery by the specified date. By reducing the lead time, the order can be delivered faster, and by anticipating the lead time when a client places an order, the manufacturer can enhance the likelihood that the product will arrive on-time.

MTS environments employ standard components and can calculate an average lead time for each product group, and the lead time is typically the same for all goods in the same group. In contrast, the lead time will be different for each order in an MTO environment due to the high customization, and it can therefore not be obtained in the same way. Lead time estimation in MTO companies is today usually based on intuition and experience, an average mean, or manual job-by-job predictions (Benjaoran and Dawood 2005; Mourtzis, Doukas et al. 2014; Schneckenreither, Haeussler et al. 2021). However, it is challenging to establish precise projections due to the variable nature of MTO production. In the literature, a variety of methods have been utilized to estimate average lead times, focusing on rule-based approaches and quantitative modeling techniques. These methods do not take lead time variability, dynamical changes or order uniqueness into account, which are important factors in MTO companies (Lingitz, Gallina et al. 2018; Szaller, Béres et al.

2018; C. Wang and Jiang 2019).

In order to manage the complex and dynamic environment in MTO, novel strategies based on intelligent approaches have received increasing attention for lead time prediction (LTP). Wuest, Weimer et al. (2016) mentioned machine learning (ML) as a highly relevant technique, due to its ability to handle dynamic, complex systems with high dimensional data and non-linear relationships. Mourtzis, Doukas et al. (2014) argued that ML models are the most robust methods for LTP. The goal of ML is to build a model based on observations and recorded data. The model can further be used to detect patterns and implicit relationships in a given dataset (Haynes, Helms et al. 1991; Pham and Afify 2005; Samie, Bauer et al. 2019; Wuest, Weimer et al. 2016). The model itself is merely one component in order to successfully develop a ML model. The Cross-industry standard process for data mining (CRISP-DM) is a well applied process model for the ML-life cycle. It includes the steps from understanding the business, gathering and preparing the data, developing and evaluating the model, and to finally deploying it (Clark 2018). It is one of the most popular and common industry-independent process models that describes approaches for different data science processes, such as ML (Clark 2018; Schröer, Kruse et al. 2021). The ML-life cycle is heavily data dependent. Technology for data collection and connected technology are becoming common to implement in manufactures, mainly through barcodes and Radio-Frequency Identification (RFID) systems (Bender, Trat et al. 2022). The availability of data increases the possibility for intelligent planning systems (Bender, Trat et al. 2022; Oluyisola 2021; Perez 2014). Companies can gather useful shop floor data through implementation of these systems, which can be used to enhance ML in manufactures (Kusiak 2017).

The research area for LTP in the literature has increased the past decade, due to new technology and increased availability of developing intelligent planning methods. The application of ML for LTP is receiving more and more attention. Despite the increased focus, several authors point out the paucity of practical implementation in companies. In addition, the appropriate ML model is context specific and a generalized ML model for LTP does not exist, meaning that each company needs to fit the model for it's purpose (Kramer, C. Wagner et al. 2020). The authors were unable to find any earlier research that considered LTP in the fenestration industry. Thus, a research gap for the use of LTP in this industry has been identified. Therefore, this thesis aims to take a first step to fill this gap.

Data collection technology increases the likelihood of using ML techniques for LTP, but the risk of collecting poor data for ML also increases since the systems might contain several flaws. Inadequate data for the task at hand or system defects making the data unusable, results in poor ML models due to the models heavily data dependability (Juddoo 2015). To overcome this, several of the methods are developed upon simulated production data (Bender and Ovtcharova 2021; Gyulai, Pfeiffer, Bergmann et al. 2018; Gyulai, Pfeiffer, Nick et al. 2018; S. Hsu and Sha 2004; Meidan, Lerner, Hassoun et al. 2009; Meidan, Lerner, Rabinowitz et al. 2011; Mezzogori, Romagnoli et al.

2019, 2021; R. Murphy, Newell et al. 2019; Schneckenreither, Haeussler et al. 2021). However, simulated data simplifies crucial manufacturing-related issues and could lead to poor model performance in the real-world. Deficits in using real-world data as the foundation for training LTP models has also been noted in the literature assessment of Burggräf, J. Wagner et al. (2020). RFID systems, as noted, is a method for gathering real-world production data. There are only detected a few articles that explicitly state that RFID shop floor is used as a data source for ML-based LTP (Fang, Guo et al. 2020; C. Wang and Jiang 2019), but these consider deep learning approaches. Data preparation, which converts raw data into useful ML input, is a crucial component that becomes more apparent when using real-world data (D. C. Corrales, Ledezma et al. 2018; Gudivada, Apon et al. 2017; Juddoo 2015). Many issues may arise while using RFID data for LTP, including sensor nodes malfunctions, calibration issues, poor sensor nodes quality, environmental effects, external noise, and networks or communication errors (Alwan, Ciupala et al. 2022). To the best of the authors knowledge there has not been conducted an in-depth evaluation of data quality for ML-based LTP. The use of RFID for the aforementioned task is neither explored within the fenestration industry.

The use of ML in a real-life problem requires cooperation from both data scientists and production managers, in order for all parties to gain the necessary knowledge. Lack of understanding the company's needs could cause the company objectives to be incorrectly translated into ML issues (Meidan, Lerner, Hassoun et al. 2009). In this thesis the authors has therefore been closely co-operating with production managers from an MTO fenestration company, Gilje, and interviewed two additional companies within the same industry. Several articles explore various LTP models, however the majority of them do not go into detail on the applications of the models within PPC. Thus, companies may not clearly see the advantage of deploying such an LTP model, which could further raise the bar for real-world application. Production managers' lack of interest may prevent the data scientists from getting the necessary data or input from actual businesses to fully comprehend the issue. Inputs from fenestration companies is especially needed, since the use of LTP within this industry has not been properly explored yet.

1.2 Research objective and Research questions

The challenges and issues mentioned in the preceding section highlight a research gap with relation to integrating lead time prediction in PPC for the fenestration industry. The thesis seeks to reflect a shared understanding between data scientists and production managers, by constructing a process model that may guide the implementation of ML-based LTP for MTO fenestration businesses. The overall scientific goal of this study is therefore to:

- Contribute to increased understanding of how ML-based lead time prediction can be developed and integrated to support PPC in the MTO fenestration industry.

In order to attain this goal, three research questions (RQs) and coherent objectives has been defined to guide the research. The RQs and objectives are summarized in Table 1.

RQ 1: *In what areas of production planning and control can lead time prediction be applied in MTO fenestration companies?*

RQ 1 aims to detect which areas of PPC the literature point out to benefit from LTP. Furthermore, these PPC areas are considered in the scope of MTO fenestration companies. Therefore, the objective is to identify PPC tasks that can benefit from LTP based on previous research, and examine whether the identified tasks can be applied within the fenestration industry through interviews of three Norwegian fenestration companies in Norway. The RQ is consistent with the main objective by assessing the deployment potential for ML-based LTP and highlight how it can provide business value for the companies.

RQ 2: *What can be the causes for data quality issues in data applied for lead time prediction?*

The research questions aims at detecting three elements of data quality. Firstly, it suggest to examine data quality issues related to collected data for LTP. Thereafter, potential root causes for the issues and potential improvement strategies for companies to overcome them are discussed. Therefore, the objective is to conduct a data quality assessment to identify data quality issues, root causes and improvement potential for the collected data. Literature review about causes for poor data quality, and assessment of the data provided by the case company will be used to achieve this objective. RQ 2 aids in the achievement of the overall scientific goal of this study by increasing the data understanding, and detecting data that needs to be eliminated before being used as input for the ML model.

RQ 3: *How can CRISP-DM be applied to predict lead times in the MTO fenestration industry?*

The last RQ aims to connect the elements of a ML-life cycle be employing the process model CRISP-DM. The results form answering the first two RQs are integrated in the answer of RQ3. The RQ implies to identify business characteristics in the fenestration industry. Thereafter, understanding, and preparing data for ML input. Data understanding is achieved through RQ2 and the business understanding, which creates the foundation for data preparation. Thirdly, ML models are developed and evaluated. This includes exploration of ML models used for LTP in the literature, and putting the desired models into practice. Finally, the application areas of PPC for LTP, detected in RQ1, shows the deployment potential for ML-based LTP in the fenestration company. The objective of RQ3 is therefore to extend CRISP-DM to a process model for LTP in the MTO fenestration industry. The case company, which operates in the MTO fenestration industry in Norway, will be used to achieve RQ3. This RQ contributes to the actual development of ML models for LTP with the use of real-world RFID data.

Research questions	Objectives
RQ1: In what areas of production planning and control can lead time prediction be applied in MTO fenestration companies?	Identify PPC tasks that can benefit from LTP in previous research, and examine whether the identified tasks can be applied within the fenestration industry
RQ2: What can be the causes for data quality issues in data applied for lead time prediction?	Conduct a data quality assessment to identify data quality issues, root causes and improvement potential for the collected data
RQ3: How can CRISP-DM be applied to predict lead times in the MTO fenestration industry?	Extend CRISP-DM to a process model for LTP in the MTO fenestration industry.

Table 1: Summary of research questions and objectives

1.3 Research Scope

Products and manufacturing strategy

The focus of this thesis is on manufacturing firms that create products tailored to individual customers, leading to high product variations and low production volumes for each product variety. Therefore, MTO is the manufacturing strategy that is the focus of this thesis. Relevant literature from research that investigates MTS and Engineer to order (ETO), which is a customer-based strategy that includes the design process as well when a customer order is placed, has also been reviewed, since there are approaches that can be derived to fit MTO environments. The sector under consideration is the fenestration sector, which creates windows and doors. All industrial categories are included in the LTP literature review since no literature was discovered examining LTP techniques utilized in this sector. Companies in the Norwegian fenestration industry make up the evaluated case companies. However, the developed approach may also be of relevance for fenestration companies in other nations that operate in comparable contexts.

PPC tasks

Three planning levels make up the standard production planning hierarchy: strategic, tactical, and operational planning, ranging form long- to short-term horizon (Kapulin and Russkikh 2020). Operational planning range over the horizon where there are actual customer orders, thereby including the order fulfillment process (Oldebråten 2017). The order fulfillment process includes actions relating to product design, production, delivery, and reporting of order status, which are of specific interest in MTO companies, since these planning tasks takes place after an order is accepted (Cox and Blackstone 2002). Different PPC tasks that are conducted during the order fulfillment process are the main topic of this thesis.

Lead time

This thesis defines the lead time as the time between the entry of a component into the shop floor and its placement in the finished goods inventory. This thesis excludes the time between the order placement and the commencement of production, as well as the time from the order is placed at the finished goods inventory until the client has received the product. This scope is established because the used RFID tracking only includes shop floor records for the designated time period and lead time.

Lead time prediction

There are numerous approaches for LTP, including machine learning, quantitative modelling, and rule-based approaches (Burggräf, J. Wagner et al. 2020; Öztürk, Kayalıgil et al. 2006). This thesis focus on ML techniques, and mainly supervised regression models. The motivation for using supervised regression models can be found in a prior research conducted by the authors as part of their specialization project in the fall of 2022, during which they reviewed the literature on LTP techniques and came to the conclusion that supervised ML models comprised the most promising ones.

Data for lead time prediction

There are several data sources that can provide valuable data for the task of LTP. Since the scope of this thesis is ML-based LTP, the data scope is narrowed down to data suitable for ML input. A literature gap was detected for the use of real-world RFID data, which will be the main focus in this thesis. In addition, this is one of the most broadly used sensor systems in manufactures.

1.4 Thesis outline

The thesis outline is presented in two ways. Firstly, the structure and a brief outline of each section is presented in Table 2. Thereafter, the whole research framework can be viewed in Figure 1. This summarizes the utilized approached in the thesis. Three main topics were viewed in the literature, which formed three research questions. The applied research methods has contributed to answering the research questions. The methods and answers from RQ1 and RQ2 contribute to the outcome of RQ3.

Section	Content
Section 1 **Introduction**	Introduces the problem background of this master thesis. Research objectives, research questions and research scope is presented, as well the thesis outline.
Section 2 **Methodology**	Presents and justifies the utilized research methods, which consist of literature reviews, case study and multi-company interviews, and a data analysis.
Section 3 **Theoretical background**	The necessary theory for the thesis is presented and defined. In addition, the conducted literature review that forms the theoretical background for the thesis is presented. This includes review of literature related to PPC application areas for LTP, ML approaches for LTP, and data quality.
Section 4 **Case study**	Presents the case company that has provided this study with data and insight in the fenestration industry. The company's production planning and production process is outlined, as well as their use of RFID data.
Section 5 **Data analysis**	The conducted data collection and data preparation are explained. As well, the data modelling is included, which concerns the utilized tools for building the ML models.
Section 6 **Results**	The results are three folded. Firstly, the results from the interviews with fenestration companies about PPC application areas for LTP is presented. Thereafter the results from the data quality assessment is presented. Finally, the results from the ML experiments are visualized and explained.
Section 7 **Discussion**	Discuss the findings from Results, and compare them to the theoretical findings from the literature review for each RQ. Based on the results, a process model for ML-based LTP that may be employed by companies within the fenestration sector is presented.
Section 8 **Conclusion**	Summarises the main findings and contribution from the master thesis, and suggestions for future research.

Table 2: The thesis structure.

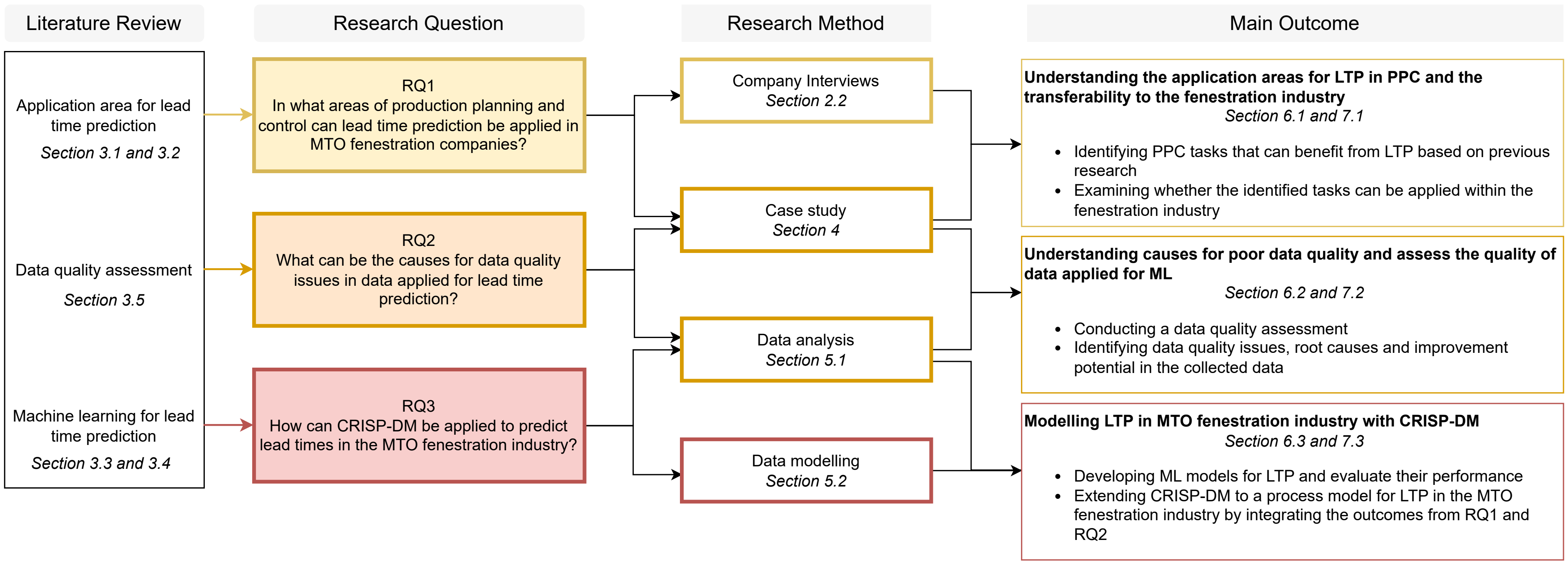


Figure 1: The thesis' research framework

2 Methodology

In this section the methodologies applied in the study are presented. The method is composed of four parts; a literature review, interviews of three companies that fit the thesis' scope, a case study and a data analysis. The companies used for the interviews are three Norwegian MTO companies, Gilje, Nordic Door and MagnorVinduet, within the fenestration industry. A more thorough examination of Gilje was then performed through a case study. Production data from Gilje was utilized in the data analysis to build a ML model that predicts lead times.

The thesis considers both quantitative and qualitative methods. Quantitative research is applicable to problems that may be described in terms of numerical and non-descriptive data which can be analyzed using statistical and mathematical methodologies, and it is based on the measurement of a given quantity or amount (Rajasekar, Philominathan et al. 2006; Sibanda 2009). To gain an accurate evaluation and statistical analysis, the data collected for quantitative research must adhere to the same rigid and consistent structure. Comparatively, qualitative research focuses on problems concerning non-numerical and descriptive data, which can be applied to gain deeper insight into a topic (Pathak, Jena et al. 2013; Rajasekar, Philominathan et al. 2006). In-depth interviews and discussions are often applied in order to collect qualitative data, and the data can further be used to get an understanding of how and what people think about a certain topic (Rajasekar, Philominathan et al. 2006).

2.1 Literature review

Several LRs were conducted in order to map and evaluate relevant concepts surrounding the application of lead time prediction, ML and data quality.

The literature review on lead time application are presented in section 3. Findings from the literature review were used to gain theoretical background and enough insight to answer RQ1 and RQ3 presented in section 1.2. Hiller, Deipenwisch et al. (2022) performed a literature review to determine the potential applications for lead time prediction. 24 papers considering lead time prediction were included in the LR. This thesis has further examined the publications included in the study and draws inspiration from the work conducted by Hiller, Deipenwisch et al. (2022). In addition to the articles found by Hiller, Deipenwisch et al. (2022), 21 additional articles were added to the literature review based on a new literature search.

The search engines, Google Scholar and Scopus were mainly used to identify the relevant litterateur. Different search terms were used to identify papers pertaining to the use of lead time prediction within MTO companies. The search terms included "lead time" and other frequently used synonyms, such as " cycle time ","processing time " and "throughput time ". In order to find publications relevant to the thesis's defined manufacturing environment, these were coupled with search terms like “make to order” and “manufacturing industry”. Other search terms relating to intelligent manufacturing, such as “prediction”, “machine learning”, “regression”, and “artificial intelligence” were also added to reduce the number of articles. The findings from the literature review were used to develop a framework that maps different application areas for lead time prediction, see section 3.2, and to examine previous ML approaches for lead time prediction, see section 3.4.

The same search engines were applied for the additional LRs. The literature review about ML included search terms like “machine learning”, “artificial intelligence” and “deep learning”, coupled with the relevant ML models for LTP, detected from the aforementioned LR, “random forest”, “artificial neural network”, and “perceptron model”. A third search block was included with “regression” and “regressor” to limit the scope. Lastly, the literature review for data quality was derived from the search terms “data quality” “data quality dimension”, and “data quality assessment”, and similar synonyms. These search terms were in turn combined with the three topics of ML and big data, internet of things and sensor systems, and production planning and control and organizations, as presented in section 3.5.

All search was limited to articles released after 2000. Additionally, the search was restricted to academic journals, conference proceedings, and books. A reading of the papers’ abstracts was completed once the number of articles was reduced. The papers that were relevant were read in full, and the most pertinent articles underwent backward and forward citation search. Instead of collecting and analyzing quantitative data, the literature review focus on a qualitative analysis, interpretation and review of the collected information from the articles. As it entails reading and evaluating existing literature to develop a thorough understanding of a specific topic, the literature review might be viewed as a qualitative approach.

2.2 Multi-company interviews

The authors conducted an interview with three companies withing the fenestration industry, namely Gilje, MagnorVinduet and Nordic Door. The objective of the interviews was to map which of the application areas identified in section 3.2.2 the companies saw as most relevant for lead time prediction. During the interviews, representatives from the companies were asked to rate the application area based on relevancy and potential. The rating ranged from low-, moderate- or high relevance. Company characteristics and current production system were mapped out during the first part of the interview for both Nordic Door and MagnorVinduet. For Gilje this was conducted by the authors in a prior case study of Gilje.

Gilje

The interview with Gilje was conducted with several of the employees, including IT and technical managers at the company. Gilje Tre AS (Gilje) is a Norwegian manufacturing company that produces highly customized doors and windows in the fenestration industry and operate in an MTO environment. They have a yearly production of approximately 120,000 windows and 20,000 doors, and there are 255 employees working at the company. This thesis will only focus on their window production, where Gilje has approximately 10% of the market share in Norway.

Gilje produces products with a wide variety of sizes, shapes, and colors, making it challenging to standardize the production process. However, Gilje has implemented an online window system that enables customers to create customized windows based on standard products. Customers can alter the window's size, features, and accessories to get the exact window they desire. Fixed casement windows are the most popular type of window. Customers often place an order a few weeks or months before the order is set to be finished.

Gilje is continuously investing in new technology to enhance the customized mass production efficiency. They have established and adhered to it's manufacturing system based on lean principles for a considerable period of time, which is a philosophy that aims to remove waste in the manufacture. Which has shown improvement in various performance indicators. They state that a lot of the non-value adding time occurs in transmitting data between different data systems. Additionally, they have implemented an RFID system in their production line to capture real-time production data. The data is partially utilized at the shop floor, but there exists several unexplored opportunities to utilize the data further for PPC. Gilje has also integrated a CalWin ERP system, which is commonly used in the fenestration industry. The ERP system stores valuable information about the orders.

Gilje has a line layout at their shop floor, where all components go through the same route. Due to implementation of lean practices, Gilje has obtained a relatively high flow efficiency by producing components as if they have the same needs. Most of the production is fully automated through conveyor belts, however they use some trolleys in order to transport the components between the different parts of the production. Additionally, the last part of production is manual, during which the workers manually assemble the products.

Gilje's lead times range from one to two days, and it has even been as low as six hours. The lead times at Gilje are affected by several factors, which include both system status and product characteristics, explained in detail in section 4.5.1. Gilje specifically mentioned bottleneck as one of the factors that has a significant impact on the lead time. The bottleneck causes long waiting times for components at the shop floor, which in turn significantly affects the lead time. Other product characteristics, such as length and product type, is also mentioned as influential factors.

Gilje does not use lead times actively in their production planning. The capacity at the fabric is solely based on the amount of glass panes in a window. Customer due dates are typically set by the customer, however if Gilje figures out that there is insufficient capacity at the given date, a negotiation is conducted between the company and the customer. After the order is accepted, it is scheduled for production. Gilje has employed backward scheduling to determine the start of production. The components that are to be made a given day follow a first-in-first-out principle based on the order placement date.

Nordic Door

The interview with Nordic Door was conducted with the IT manager at the company. Nordic Door is a Norwegian door manufacturing company with approximately 180 employees. The company provides highly customized doors in the MTO fenestration sector, and the produced doors may differ greatly in color, size, locking system and material, however a standard office door is the product that they produce most frequently. The company claimed that because they have such a wide variety of products it is challenging to categorize them. Therefore, they do not have any product groups, meaning that they don't categorize their products into distinct groups based on product characteristics. Due to this high customization, the customers of Nordic Door are not able to configure a door over the website based on standard products like Gilje does. In the current system, an architect sends a sketch to Nordic Door with requirements for the door. This communication is conducted by email. How far in time a consumer places an order might vary greatly, but typically it is around 2 months.

As opposite to Gilje, Nordic Door haven't implemented any data systems that are able to capture data along the production line, such as the RFID system. The company's data is similar to Gilje, stored in a CalWin ERP system. The system provides the company with information about how long it takes for an order to be delivered to the customer after it has been released to the ERP system.

The production layout is described by Nordic Doors as a "good old-fashioned industrial company". The company operates with a functional layout, where each product has a unique routing. Nordic Door stated that the layout is not as streamlined as they would have liked, and that they use a lot of trucks and conveyors in order to transport the components between the machines. As mentioned, the routing for each unique component does also vary, adding to the system's complexity. There are both modern and old machines in the fabric, but overall there is relatively low degree of automatization.

The Nordic Door representative stated that the lead time was approximately 14 days, but it was difficult to be more precise because there was no specific control or measurement of the lead time in their manufacturing as of today. Additionally, they don't alter the lead time between different products. He added that although the lead time has been as short as 10 days, it might also be much longer than 14 days due to several manufacturing halts in their production line. The lead times at the factory may be affected by several factors, such as bottlenecks that can cause stoppages, raw material delays, and order characteristics that may have an impact on the product routing. The company has placed a buffer storage between almost every machine since the production line suffer from a lot of waiting time. Numerous orders must be re-prioritized in order to meet the delivery date due to the frequent production delays, and several times these delays force Nordic Door to be entirely unable to fulfill the delivery date.

The current production planning at Nordic Door does, similarly to Gilje, not place much weight on the lead time. Regarding their capacity planning, they use the amount of doors as a capacity measurement. The total amount of doors that the company are able to produce in a day is determined by the capacity of the system's bottleneck. If they have enough capacity, the order is accepted and the customer due date is set according to the customer's preference. The different types of doors have quite varying lead times, however this is not taken into account when calculating capacity. Regarding their production scheduling, Nordic Door has shifted from weekly to daily production scheduling, focusing each day on production optimization by examining the sorting of the orders that are to be made that day. The objective is to fulfill all of the orders scheduled for the given day. For this transition to work efficiently, minimized machine setup and changeover times are of need, which the company is still working with. The pace of the operators has a big effect on the machine setup time and the changeover times.

MagnorVinduet

The factory coordinator at MagnorVinduet was their representative during the interview. MagnorVinduet is a window and door manufacturer from Norway operating in the MTO fenestration industry. There are around 65 employees working at MagnorVinduet as of today. The company produce a wide range of customized windows and doors, however this thesis will only focus on the manufacture of windows. Similar to Gilje and Nordic Door, all goods are produced according to the specifications of the customer and may vary in size, color and material. The window type they make the most of is the "H-window", which is characterized by its horizontal orientation ability. MagnorVinduet state that no windows are alike, and that every week is unique as a result of the great level of customisation and variation. The ordering of new windows are conducted over the phone or via email. The window specifications are established through this conversation, and the seller will thereafter presents the customer with an offer, containing the price and delivery date. Orders are frequently placed weeks or months in advance of the delivery date, and the company has a minimum delivery time of four weeks, due to supplier lead time of three weeks and production time of one week.

Regarding data management, MagnorVinduet applies a self-developed Excel system in order to handle incoming orders and to conduct the production planning. As of now, they have no system that are able to capture data along the production line. MagnorVinduet stated that they previously adopted an RFID system. The company however, decided to discontinue usage of it since it led to a large amount of defective manufacture because of a great deal of problems related to the RFID tags, such as dysfunctional tags from the supplier, and the tags falling of.

The production system at MagnorVinduet can roughly be divided into four work stations; a wood station, an aluminum station, a glass station and an assembly station. There is some distance between each of the stations, and various means of transportation, including trolleys and trucks, are frequently employed in order to transport the products between them. Within a work station the layout may resemble a linear layout, however the overall layout is rather functional. The routing of a products is specific to that given product and depends on its needs, such as whether it needs aluminum cladding. The production system is characterized by a lot of manual work, and the representative from MagnorVinduet stated that almost 70 % of the production was dependent on manual work. A large raw material storage is not required, since the majority of raw materials are bought after an order is placed. However, they have several buffer storages between each work station since the production is influenced by a lot of waiting time.

MagnorVinduet struggled to provide a precise estimate for the lead times at the factory. However, they stated that the minimum lead time is estimated to be two days, and for the production planning they always assume a standard lead time of around a week due to their current planning procedure which spans over a week. There are several factors that affect the lead time at MagnorVinduet. One of the most influential factors mentioned by the company is the impregnation of the wood material. After impregnation, the material has to dry for 24 hours, further increasing the total lead time. Bottlenecks that occur due to exceeded machine capacity were also highlighted as a lead time influencing factor. Other mentioned factors included setup time, delays due to manual labour and order characteristics. The company claimed that there were no significant concerns with production or delivery delays, and that any delays that did occur were likely caused by issues with raw materials, such as delays at the supplier or defective goods.

Similar as for the two other companies, neither MagnorVinduet use lead times in their planning. The due dates at the company may be determined by the customer and the company. The company suggest a delivery date based on available capacity or the customer can specify a delivery date. Regarding their capacity planning, the company employs the developed data system mentioned above. When the order is added to the system, a built-in function calculates the cost and time estimates for the given order including the time period from the purchase of raw material until the whole order is completed. These calculations are based on fictitious numbers, experience and some research. The data is updated frequently to reflect the most recent trends. In order to evaluate capacity, the production manager strives to strike a balance between time and cost considerations. Experience is also taken into account when determining the capacity. If the company finds that capacity is being exceeded, they may work overtime to prevent production delays. MagnorVinduet follows a weekly production schedule, where all products belonging to an order scheduled for that week are sorted according to the type of product. They begin the following week's production if they complete all of the orders scheduled for the current week, and they adjust their plans for the following week if they are unable to complete all of the orders. If the customer places an order several months in advance, the company usually produce the order as soon as they have available capacity, and place it at the finished storage until the date of the delivery.

2.3 Case Study

A case study, according to Gerring (2006), is a thorough examination of a single unit or a small number of units with the aim of understanding a larger class of similar units. Researchers can gain extensive and important insight into the characteristics of actual world occurrences through case studies (Yin 2009). It can be challenging to distinguish between characteristics that are unique to the unit under investigation and those that can be utilized to draw generalizations applicable to all companies withing the same scope (Gerring 2004).

The case study is a continuation of previous work conducted by the authors in their project thesis the fall 2022. The case company used in this thesis is as mentioned Gilje. The case study was carried out in order to gain practical knowledge and information from a real-world company that fits into the scope. Information was gathered using qualitative techniques such as in-depth interviews, observations through a visit at their factory and discussions with the case company. By employing these techniques, qualitative data was collected, and the authors were able to map the variables at Gilje that affect the lead time with the use of this information. It also provided valuable insight into Gilje's production system and how they collect data using sensor technology. The obtained knowledge was further used to decide which factors to include in the data analysis and from which data sources.

2.4 Data analysis

Data analysis and modelling was performed in order to predict lead times with the obtained data from Gilje using ML. The authors adhered to the procedures in the the process model CRISP-DM, which includes the steps that are executed both before and after the actual development of the ML model, and it can be broken down into six separate steps; (1) business understanding, (2) data understanding, (3) data preparation, (4) modeling, (5) evaluation and (6) deployment.

(1) Business understanding: The initial phase is carried out to comprehend the company characteristics, company constraints and to frame the issue. This makes it easier to select appropriate input features and ML models, as well as evaluating the results. In this phase, the determination of the project objectives, as well as identification of company needs and constraints are carried out in order to achieve an in-depth understanding of the business problem and the success criteria for the project (Schröer, Kruse et al. 2021). In a study conducted by Schröer, Kruse et al. (2021), case study was identified as an approach previously used for the business understanding phase. In this thesis, Gilje is thoroughly examined through a case study, see section 4. This provides the authors with useful information that can be utilized in order to comprehend the problem and determine

which aspects to take into account when predicting the lead time in a fenestration company. Inputs from Nordic Door and MagnorVinduet were also taken into account, together with the findings from the LR, in order to generalize the characteristics of the industry.

(2) Data understanding: Understanding the data is the second stage. Since ML models are heavily data dependent, having appropriate data is essential for producing a model that can provide accurate results. In this step, pertinent data sources must be found in order to obtain features that, according to the business understanding, should be included. An evaluation of the data should further be performed to ensure the quality of the data, and the results could be used to choose which data to include in the final dataset. Previous approaches used to explain the collected data includes; descriptive statistics, data visualization, showing examples of the data and conducting interviews with experts (Schröer, Kruse et al. 2021). Results from section 5.1.1 shows how the data was collected and from which data sources. Section 5.1.2 present the data preparation techniques that was applied, and section 6.2 provides an understanding of the quality of the collected data through statistical analysis of the data.

(3) Data preparation: Data preparation is the third step of the CRISP-DM model, and is one of the most time consuming parts of the ML process. In this step the collected data is cleaned in order to remove data quality issues. Thereafter the data is pre-processed and transformed into an understandable input that can be used by a ML model. Cleaning the data and achieving a complete dataset to use for the ML model is crucial in order to get reliable results. In general, this phase includes the steps of removing or replacing missing values, duplicate data, invalid data and noise (Javatpoint 2023; Maharana, Mondal et al. 2022). To ensure that every feature are in the same range, the data is commonly scaled. Changing from categorical values to numerical values using different encoding methods is also a common practice in this phase. Maharana, Mondal et al. (2022) stated that encoding is the most significant factor, affecting supervised ML models ability to achieve great performance. The data preparation techniques used in this study are described in section 5.1.2.

(4) Modeling: Even though the modeling phase is one of the most crucial procedures, it is also one of the shortest. The suitable ML models are chosen based on the business problem and the available data (Schröer, Kruse et al. 2021). The chosen models for this study is based on the findings from the LR, presented in section 3.4 Before applying the data as inputs in the model, it is split into a training and testing set (Clark 2018). The models are further trained utilizing the training data obtained in the preceding steps, and it can be beneficial to train more than one model to increase the number of evaluation foundations (Schröer, Kruse et al. 2021). The model is able to learn different patterns, rules, and characteristics based on the training dataset. The test dataset is used to evaluate the model. In order to discover the ideal collection of hyperparameter values that can be used to improve the performance of the model, different hyperparameter tuning techniques are employed, such as grid search and randomized search. The modeling process is

described in section 5.2. The authors decided to use two different ML models, namely MLP and RF, due to their popularity and superior performance in prior studies.

(5)Evaluation: This step focuses on assessing the performance of the model using different evaluation criteria or metrics, which is also highlighted by Schröer, Kruse et al. (2021) as a common approach in order to evaluate the quality of the models. Through a proper evaluation, the accuracy of the model is validated. The test dataset obtained in the modeling phase is used as input in the ML model in order to asses and evaluate the accuracy of the ML results, and check whether the project objectives developed in the business understanding phase are met. The evaluation metrics used in this thesis are described in section 5.2.1, and the results are displayed in section 6.3. The results consist of three different experiments. The first experiment examines and contrasts the model's performance when using a dataset that includes parameters from the system status and when using a dataset that does not. Experiment 2 and 3 includes a hyperparametertuning for the two different ML models.

(6)Deployment: The final phase of the ML process is the deployment. In this phase, the model may be deployed into a real-world system. Both a deployment plan and a monitoring and maintenance plan should be implemented at this stage. According to Schröer, Kruse et al. (2021), only a few research take the deployment phase into account, and this study will not deploy the developed ML model either. However, the authors tried to narrow the gap by identifying relevant application areas where a lead time prediction model potentially could be employed. The identified areas are discussed in section 7.1, and the provided information could be used for future deployment plans.

3 Theoretical background

In this section the theoretical background and the results from the literature reviews will be introduced. Relevant concepts and terminology related to the manufacturing environment are described. Thereafter the literature review on lead time prediction is presented, along with the identified application areas for LTP in PPC, followed by concepts related to ML, and literature assessing ML-based LTP. Lastly, the reviewed literature about data quality is presented.

3.1 Manufacturing environment

Companies operate in various manufacturing environments, determined by the products they produce and the products manufacturing needs. An understanding of the manufacturing environment is fundamental in order to apply the proper planning techniques (Jonsson and Mattsson 2003).

The order decoupling point (ODCP) is a point in the manufacturing value chain for a product, where production based on actual customer order is separated from production based on forecasts and plans (Buer, Strandhagen et al. 2018). A company's placement of ODCP is used to categorize it's manufacturing environment, as presented in Figure 2. Forecasts trigger the production operations upstream of the ODCP, while customer orders trigger the actions downstream of the ODCP. The common categorization of manufacturing environment are make to stock, assemble to order (ATO), make to order and engineer to order, listed from less costumer involvement to most (Buer, Strandhagen et al. 2018; Chapman 2006; Sagegg 2020).

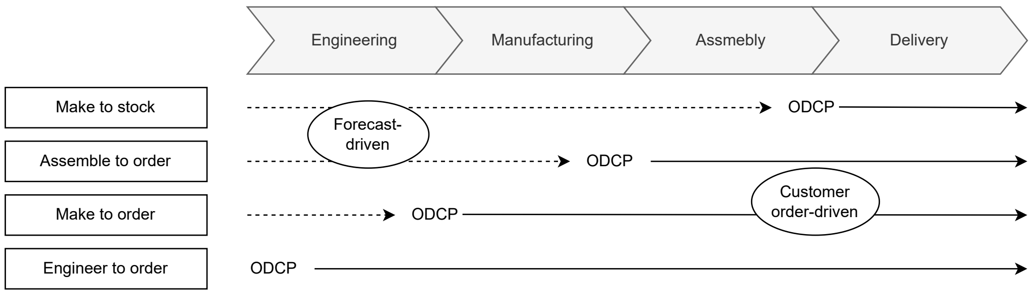


Figure 2: Manufacturing environments described by the order decoupling point

MTO manufacturing environments is the focus of this thesis. MTO strategies are customer based, and the manufacture starts after a customer order is placed (Chapman 2006; Sagegg 2020). This results in volatile demand which rarely can be predicted (Linda C Hendry and Kingsman 1989). MTO strategies provides highly customized products, which are commonly produced in small batches (Oldebråten 2017). The majority of MTO companies base their products on a mix of standard components and customer-specific components (Linda C Hendry and Kingsman 1989; Sagegg 2020). The planning environment is significantly more dynamic and uncertain, compared to MTS and ATO, due to MTO companies' usual customer engagement (Thürer, Stevenson et al.

2014).

3.1.1 Production planning and control in MTO

Production planning and control is frequently referred to as the necessary actions to balance supply and demand (Buer, Strandhagen et al. 2018 derived from Vollmann, Berry et al. 2004). PPC involves the crucial tasks of organizing, planning and performing manufacturing activities (Cañas, Mula et al. 2022). In order to meet customer demand, PPC systems aid decision-makers in deciding what, when, and how much to produce, purchase, and deliver (Cañas, Mula et al. 2022). The planning and decision-making processes used by the manufacturer have a big impact on the company's performance (Tenhiälä 2011). The success or failure of a manufacture has also been noted to be significantly influenced by effective PPC (Buer, Strandhagen et al. 2018 derived from Vollmann, Berry et al. 2004).

The appropriate PPC techniques for a manufacture differ depending on the manufacturing environment and sector. The ODCP's position marks a turning point for the need of various strategies and techniques for PPC, downstream and upstream of this point. The traditional production planning hierarchy typically contains three planning levels; strategic-, tactical- and operational planning (Hung, C.-C. Huang et al. 2013). Depending on the level of detail and the planning horizon, different decision-making tasks are classified under one of the three planning levels (Kapulin and Russkikh 2020; Oldebråten 2017). Strategic long-term planning determines the general directions, structure and goals for the company over a time horizon of several years (Kapulin and Russkikh 2020; Muñoz, Capón et al. 2012). Decisions made at this stage are determined by long-term forecasts (Kapulin and Russkikh 2020). At the tactical level a production plan is derived from the company's strategies defined at the strategical level. The plan should balance market demand and the manufacture's capacity (Kapulin and Russkikh 2020). Tasks that frequently relate to the tactical level include Master production schedule (MPS), Capacity requirement planning (CRP) and Material requirements planning (MRP). The MPS reflect the planned production for the company, the CRP involves planning the production capacity that are needed for the MPS, and MRP ensures that the right component is in the production at the right time (Kapulin and Russkikh 2020). Activities at the tactical level, including MPS, CRP and MRP, are hard to obtain in an MTO environment, since the product parameters are unknown until an order is placed. Where in contrast an MTS company can follow this procedure since they produce based on forecasts and the product parameters are known in advance (Kapulin and Russkikh 2020). The operational planning level is the last level of the planning hierarchy. It relates to short-term planning and scheduling of orders on a daily or weekly basis (Kapulin and Russkikh 2020). In this phase, tasks are assigned to different resources, and the order sequence for each resource is decided through a more detailed production plan (Kapulin and Russkikh 2020; Muñoz, Capón et al. 2012). The plans created at the operational level have a planning horizon that corresponds to the period of time where there are actual customer orders, whereas tactical plans cover a time frame that spans beyond actual orders (Oldebråten 2017). Stevenson, Linda C Hendry et al. (2005) highlighted that the choice of appropriate PPC techniques is specially important for MTO companies, because of the uncertain environment they operate in.

Kapulin and Russkikh (2020) stated that the production planning for MTO companies should start after an order is received, i.e. it should start at the customer enquiry stage. According to Hiller, Deipenwisch et al. (2022), the customer enquiry stage is the initial phase of the order fulfillment process, which includes actions relating to product design, production, delivery, and reporting of order status (Cox and Blackstone 2002). During order fulfilment, several PPC tasks are performed with the goal of optimizing the production performance and making sure that that the order is fulfilled before the given due date (Hiller, Deipenwisch et al. 2022). The customer enquiry stage is therefore an specially important planning stage for MTO companies, since the production cannot start until a customer places an order. Additionally, all subsequent stages are affected by the customized product that is placed by the customer (L. Hendry and Kingsman 1993; Kapulin and Russkikh 2020). Due date setting and pricing are typical actions to be made during this phase (L. Hendry and Kingsman 1993). Capacity planning is another essential task mentioned by L. Hendry and Kingsman (1993), and by performing capacity planning at this stage the company may be able to set reliable due dates that can be achieved with the available capacity (L. Hendry and Kingsman 1993). After the customer enquiry stage, the company must plan and purchase the required materials, as well as schedule the order by determining a start and end date for the production (Hiller, Deipenwisch et al. 2022; Kapulin and Russkikh 2020). Throughout the order fulfilment process, adjustments to the capacity plan may also be performed to ensure that jobs are completed before their due date (Linda C Hendry and Kingsman 1989).

To summarize, PPC involves organizing, planning, and executing manufacturing activities to meet customer demand. It plays a critical role in balancing supply and demand by determining what, when, and how much to produce, purchase, and deliver. The manufacturing environment affects which PPC strategies are effective. For MTO companies, the production planning process often begins at the customer enquiry stage, where due date setting, pricing, and capacity planning are essential tasks. Throughout the order fulfillment process, which includes product design, production, delivery, and order status reporting, different PPC tasks are performed in order to optimize production performance and ensure timely order fulfillment by planning and purchasing necessary materials, scheduling production, and adjusting the capacity plan as needed. These tasks are considered detailed operational-level order fulfillment activities, with a planning horizon corresponding to the duration of the actual order.

3.2 Lead time

3.2.1 Lead time in manufacturing

Lead time is a key performance indicator in many PPC tasks, and short, reliable, and predictable lead times are becoming an increasingly essential criteria in order for companies to stay competitive (Asadzadeh, Azadeh et al. 2011; Fussenegger and Lange 2022; Hiller, Deipenwisch et al. 2022; Hvolby and Thorstenson 2001; Thürer, Stevenson et al. 2014). Lead time can be defined as a period of time needed to complete a process, and it may consist of multiple different components that together make up the total lead time (Cox and Blackstone 2002). From a customer perspective the lead time can be assessed as the period from a customer places an order and the order is confirmed, until the order is delivered to the customer (Kader and Akter 2014). The customer lead time can be further divided into several sub processes. This paper divides the customer lead time into three separate processes that are shown in Figure 3. The material lead time is the period of time from when the required materials are ordered until they are delivered to the factory. The time it takes to manufacture an order is known as the production lead time, and the time it takes to deliver an order from the finished goods inventory to the the customer is known as the delivery lead time. The production lead time, which will be referred to as simply lead time for the remainder of the article, is the subject of this paper.

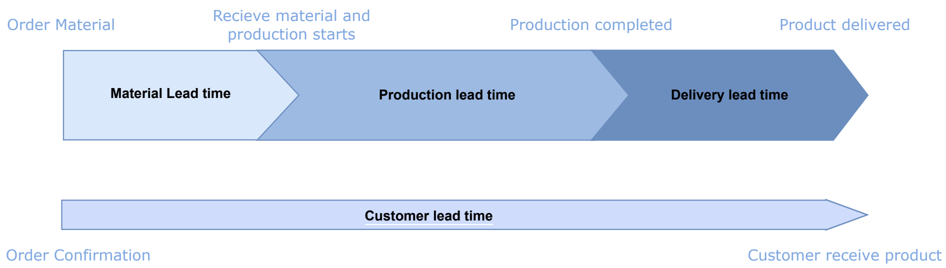


Figure 3: The different processes that make up the customer lead time

3.2.2 Application areas for lead time prediction in PPC

As previously mentioned, lead times are crucial performance indicators used in PPC, and by applying precise lead time calculations, a production plan's reliability may be improved. Lead time can also affect other performance measures such as cost and quality. Costs can be decreased by reducing lead times, however doing so may compromise product quality (Kader and Akter 2014).

There are numerous articles that examine LTP models, however many of them do not elaborate on or explain the application area of the models within the PPC hierarchy. Hiller, Deipenwisch et al. (2022) conducted a study in order to categorize previous approaches used for LTP. They identified three distinct application areas within the order fulfilment process where LTP could be of relevance; *order clarification*, *throughput scheduling* and *order coordination*. Within the order clarification phase a production plan is developed providing roughly planned lead times based on available order information. Throughput scheduling focus on short-term production planning and provides detailed lead times for each order based on relevant production data, e.g WIP, waiting time or processing steps. After the order has been released to the shop floor, tasks within order coordination are carried out with a focus on production control. The remaining lead times are determined in this phase by comparing the actual lead time to the lead times from the short-term plan using real-time data to capture the status of the system. In their literature review, Hiller, Deipenwisch et al. (2022) considered 24 articles. This thesis has further investigated the application areas by including 21 additional articles, based on a new literature search, see section 2.1.

Capacity planning

A critical phase for manufacturing companies is capacity planning, which encompasses activities related to determining the capacity requirements necessary to meet demand (Cox and Blackstone 2002). Lead times are crucial for capacity planning because they influence how long a job spends on the shop floor and occupies system capacity. Several of the articles used in this study discuss the importance of accurate LTP in order to obtain a viable capacity plan (Choueiri, Sato et al. 2020; Kramer, C. Wagner et al. 2020; Liang 2010; Meidan, Lerner, Hassoun et al. 2009; Meidan, Lerner, Rabinowitz et al. 2011). The production planners can acquire a more realistic production capacity requirement planning, which refers to the process of defining, evaluating, and revising capacity limitations or levels, by providing precise lead time estimates (Choueiri, Sato et al. 2020; Meidan, Lerner, Hassoun et al. 2009; Meidan, Lerner, Rabinowitz et al. 2011). This in turn may enhance the overall performance of the production planning. Liang (2010) mentioned lead time estimation as an essential component used in MTO environments in order to maximize and make best use of the available capacity in the production system. Kramer, C. Wagner et al. (2020) also stated the importance of precise lead times in order to determine the production system's capacities. A precise capacity plan is vital to obtain, since many PPC activities are dependent on the capacity, e.g production scheduling. Capacity planning sets the foundation for production scheduling by ensuring that the necessary resources are in place to meet production demands.

In a study conducted by Mestry, Damodaran et al. (2011), the authors created a branch and pricing solution, which is a mixed-integer linear programming (MILP) approach, for order acceptance and capacity planning in an MTO environment. The authors further stated that MTO companies must effectively manage their capacity in order to generate long-term profits and guarantee on-time delivery. Lead times are employed as constraint parameters in the MILP problem, in order to ensure that the capacity is not exceeded. Zijm and Buitenhek (1996) developed an approach for capacity planning and lead time management in an MTO environment which explicitly considered workload dependent lead times. The earliest completion date could subsequently be determined based on the lead times and utilized as an input for detailed shop floor scheduling. In MTO the lead time of an incoming order is therefore necessary to predict in order to have viable estimates to fulfill these methods.

Due date setting

Due date setting is the application area for LTP that is most frequently mentioned (Ahmarofi, Ramli et al. 2017; Backus, Janakiram et al. 2006; Gyulai, Pfeiffer, Bergmann et al. 2018; G. M. Lee and Gao 2021; Liang 2010; Meidan, Lerner, Hassoun et al. 2009; Meidan, Lerner, Rabinowitz et al. 2011; Öztürk, Kayalıgil et al. 2006; Pfeiffer, Gyulai et al. 2016; Schuh, Gützlaff, Sauermann, Kaul et al. 2020; Schuh, Gützlaff, Sauermann and Theunissen 2020; J. Wang and J. Zhang 2016; J. Wang, J. Zhang and X. Wang 2018; Alenezi, Moses et al. 2008; Gyulai, Pfeiffer, Nick et al. 2018; S. Huang, Guo et al. 2019; Kramer, C. Wagner et al. 2020; R. Murphy, Newell et al. 2019). Due date refers to the period of time between the acceptance of a customer's order and delivery of the order to the client (Cox and Blackstone 2002). This includes the amount of time between the order being confirmed and the start of production, the lead time, the time spent at the finished goods inventory and the transit time. Due date setting is often a lead time dependent task, since many due date setting approaches are based on lead time estimates. Accurate LTP can therefore be crucial for achieving on-time delivery, improve customer relations and maintain competitiveness (Alenezi, Moses et al. 2008; R. Murphy, Newell et al. 2019; C. Wang and Jiang 2019). Due dates also have an impact on scheduling and other shop floor management tasks, and reliable due date setting allows for more efficient production planning, which can help lower different production costs, such as inventory holding cost (S. Huang, Guo et al. 2019; R. Murphy, Newell et al. 2019).

Alenezi, Moses et al. (2008) stated that due date performance in MTO companies is frequently below the expectations, and in order to increase performance and cut costs associated with early or late deliveries, the company must provide accurate LTPs in real-time. Several of the articles addressed this problem by developing methods that could provide precise predictions. Gacek (2018), S. Hsu and Sha (2004) and C. Wang and Jiang (2019) developed different ML models for LTP. Results provided by the models were further used to set the due date of incoming orders. An approach using input/output control was developed in a study by L. Hendry and Kingsman (1993) to control lead times, and as a result, the delivery time of orders. The input/output control is constrained by the lead times of the system's orders. R. Zhang and Wu (2012) discussed the close connection between due date setting and production scheduling. The authors developed a scheduling optimization model that was integrated with a model for due date setting based on lead times. Mezzogori, Romagnoli et al. (2021) and Mezzogori, Romagnoli et al. (2019) developed methods for LTP using workload control. The methods' objectives were to establish due dates and maximize on-time delivery.

Production scheduling

Scheduling is also mentioned by several articles as a lead time dependent application area (Backus, Janakiram et al. 2006; Bender and Ovtcharova 2021; Bender, Trat et al. 2022; T. Chen 2009; Gyulai, Pfeiffer, Nick et al. 2018; Kramer, C. Wagner et al. 2020; Liang 2010; Meidan, Lerner, Hassoun et al. 2009; Meidan, Lerner, Rabinowitz et al. 2011; Öztürk, Kayalıgil et al. 2006; R. Zhang and Wu 2012). Production scheduling refers to the process of allocating resources to jobs in a specific order and within a given time frame, and it aims at optimizing operational tasks and processes (Parente, Figueira et al. 2020). In numerous articles, the significance of accurate LTP is emphasized for scheduling optimization and effectiveness (Bender and Ovtcharova 2021; Lim, Yusof et al. 2019; Lingitz, Gallina et al. 2018; Mori and Mahalec 2015). Many scheduling rules and techniques require lead times as inputs, and by obtaining accurate and reliable lead times, the performance and efficacy of these rules and techniques may be improved (T. Chen 2009; Gyulai, Pfeiffer, Nick et al. 2018; Lu, Ramaswamy et al. 1994; Mori and Mahalec 2015). The cost of production may also be impacted by the lead times employed in scheduling. It is important to provide schedules that meet the established production capacity, and the lead times of orders are key components in order to meet this requirement and avoid deviations that can increase the production costs (Mori and Mahalec 2015). The expenses associated with keeping inventory on hand could rise if the scheduled order is finished earlier than the specified lead time. On the other hand, the company may be charged a lateness cost if the actual lead time exceeds the promised lead time (Liang 2010). Accurate lead times that can be used to provide precise production schedules are therefor crucial in order to improve production performance while minimizing the production costs.

Lu, Ramaswamy et al. (1994) introduced two new scheduling rules called fluctuation smoothing policies for variance of cycle time and fluctuation smoothing policies for mean cycle time. Both policies are dependent on lead time, and results from the study showed a lead time reduction with an implementation of the policies. Ren, Chaw et al. (2022) employed a number of lead time-based scheduling rules to determine the optimal order sequence. In order to support managers, Cos Juez, Nieto et al. (2010), Yamashiro and Nonaka (2021) and Mori and Mahalec (2015) developed LTP models that could be incorporated into production planning and scheduling systems in order to optimize the production performance. Lead times are also a crucial factor for order release methods, according to Schneckenreither, Haeussler et al. (2021), as decisions regarding order release are frequently dependent on them. The authors developed a model for LTP using ML. Based on the resulting lead times, a planned release date was determined. Resource scheduling has also been mentioned as a lead time dependent task (Meidan, Lerner, Hassoun et al. 2009). The need for accurate LTP for order release was also discussed by Liang (2010).

Production analysis

A big part of LTP is to discover the factors affecting the lead time. Historical lead time data can be employed for this sort of descriptive analysis, in order to assess trends and patterns in the production. By identifying key factors, one can improve prediction accuracy as well as gain insight into, and enhance lead time control (Meidan, Lerner, Rabinowitz et al. 2011; J. Wang, J. Zhang and X. Wang 2018). Through production analysis of the lead time data, the cause of deviations, interruptions or lateness in the production may be discovered (Pfeiffer, Gyulai et al. 2016). According to Ahmarofi, Ramli et al. (2017), analysis of lead times has not received much attention in earlier work. The authors conducted an analysis to determine the most important factors that influences a job's tardiness based on the lead time, such as machine capacity and waiting times of the components. By identifying machine constraints, the company may be able enhance smoothness in the production, while an assessment of the waiting times can help companies improve their inventory management. Other significant factors affecting lead times that can be identified through an analysis of the lead time data include equipment unreliability, production flow, sequence-dependent setup-time, changing demand, unbalanced capacity, job reentry into machines and bottlenecks (Choueiri, Sato et al. 2020; G. M. Lee and Gao 2021). It will be possible to take proactive actions with LTP in order to cut lead times and production costs by learning about the elements affecting lead time, and in addition it could help obtain a more sustainable production (Choueiri, Sato et al. 2020; G. M. Lee and Gao 2021; Meidan, Lerner, Hassoun et al. 2009; Meidan, Lerner, Rabinowitz et al. 2011; Pfeiffer, Gyulai et al. 2016).

Control of WIP level is also an important task on the shop floor, which is highly correlated to lead time. A reduction in WIP can shorten lead times and increase production efficiency (Cao and Ji 2021; Chien, C.-Y. Hsu et al. 2012). LTP may provide companies with information that can be used to find the ideal WIP level for future orders. Chien, C.-Y. Hsu et al. (2012) developed a two-phase approach that can help production managers control and find the optimal WIP level in the production system. The approach is based on a model for LTP that could be used to control elements affecting the production status such as the WIP level. Cao and Ji (2021) outlined how accurate LTPs can be used in planning to prevent unnecessary waiting times and WIP.

Production monitoring

In the framework developed by Hiller, Deipenwisch et al. (2022), order coordination is mentioned as an application area for LTP, with focus on the remaining lead time. Tasks inside this area are applied after the order is released to the shop floor and focus mainly on production control. Continuous control and monitoring of the production system is mentioned in several of the articles as a way to improve system efficiency and customer relation (T. Chen 2009; Gyulai, Pfeiffer, Bergmann et al. 2018; Gyulai, Pfeiffer, Nick et al. 2018; Meidan, Lerner, Hassoun et al. 2009). To keep customers updated on the status of their orders and guarantee that they will be delivered on schedule, an estimate of the remaining lead time can be employed (T. Chen 2009). Meidan, Lerner, Hassoun et al. (2009) developed a model that can be used to predict only a part of the lead time, namely the waiting time, of a job. The authors further discuss the potential of using the model as a real-time decision support in order to identify potential delays in the production system and the cause of the delays. S. Huang, Guo et al. (2019) highlighted how lead time monitoring can be utilized to boost production efficiency by preventing delays, through re-prioritizing orders, stabilizing the manufacturing process, and allocating resources optimally. Gyulai, Pfeiffer, Nick et al. (2018) discussed the application of ML to predict the anticipated, remaining lead time of ongoing jobs. This would allow the chance to adjust the jobs' priorities throughout production or change the product's routing if the original route is causing the job to be late. Gyulai, Pfeiffer, Bergmann et al. (2018) explored the aforementioned method further. The proposed method can be applied in order to reduce the tardiness of an order by altering the priorities of the orders in the system in real-time.

In section 3.2.2, five application areas where identified from the literature, where LTP could be of interest. These application areas include capacity planning, due date setting, production scheduling, descriptive analysis and production monitoring. Capacity planning, due date setting and production scheduling are typical tasks performed during the production planning phase, and use the lead time as an input. LTP can be employed in capacity planning in order to determine the capacity requirements for meeting future demand. Due date setting often rely on LTP to ensure on-time delivery, improve customer relations, and optimize production planning. Several production scheduling approaches rely on accurate LTP, which can help optimize resource allocation, minimize costs, and ensure that the production capacity requirements are met. Production analysis and production monitoring are tasks within production control, which can assist with the management and optimization of production processes. LTP in production analysis can be used to test optimal production flows, and predict the impact of the changes. Production analysis involves analyzing historical lead time data to identify factors influencing lead time and use the obtained information to enhance the production performance. Production monitoring involves continuous control and monitoring of the production system using LTP to update customers on order status, prevent delays, and make changes to the current production that can improve the production flow.

As seen, accurate LTP can play a crucial role in enhancing overall production performance and efficiency across different areas of PPC. Table 3, summarizes the mentioned application areas detected from the literature and their relationship with lead time and lead time prediction.

Application area	**Definition**	**Relationship with lead time prediction**	**Sources**
Capacity planning	The determination of capacity requirements needed to fulfill future demand (Cox and Blackstone 2002).	Lead times determine the amount of time a component spends on the shop floor and occupies capacity, and it is often used as inputs for different capacity planning approaches. By integrating LTP in capacity planning, the company may be able to ensure optimal resource utilization, and obtain a more realistic and viable production plan.	*Choueiri, Sato et al. 2020; Kramer, C. Wagner et al. 2020; Liang 2010; Meidan, Lerner, Hassoun et al. 2009; Meidan, Lerner, Rabinowitz et al. 2011; Mestry, Damodaran et al. 2011; Zijm and Buitenhek 1996*
Due date setting	Determine the period of time between the acceptance of a customer's order and delivery of the order to the client (Cox and Blackstone 2002).	Due date setting is often a lead time dependent task, as the due date is typically set based on estimates of the lead time. LTP dependent due date approaches may enhance customer relation and competitiveness by ensuring reliable and realistic due dates.	*Ahmarofi, Ramli et al. 2017; Backus, Janakiram et al. 2006; Gyulai, Pfeiffer, Bergmann et al. 2018; G. M. Lee and Gao 2021; Liang 2010; Meidan, Lerner, Hassoun et al. 2009; Meidan, Lerner, Rabinowitz et al. 2011; Öztürk, Kayalıgil et al. 2006; Pfeiffer, Gyulai et al. 2016; Schuh, Gützlaff, Sauermann, Kaul et al. 2020; Schuh, Gützlaff, Sauermann and Theunissen 2020; J. Wang and J. Zhang 2016; J. Wang, J. Zhang and X. Wang 2018; Alenezi, Moses et al. 2008; Gacek 2018; Gyulai, Pfeiffer, Nick et al. 2018; S. Hsu and Sha 2004; S. Huang, Guo et al. 2019; Kramer, C. Wagner et al. 2020; Mezzogori, Romagnoli et al. 2019, 2021; R. Murphy, Newell et al. 2019; C. Wang and Jiang 2019*

Production Scheduling	The process of allocating resources to jobs in a specific order and within a given time frame (Parente, Figueira et al. 2020).	Several scheduling approaches today utilizes lead time as an important input parameter. By employing LTP, the company may obtain precise lead time estimates that can be used to develop realistic production schedules. This can further enhance the production efficiency and minimize production costs.	*Backus, Janakiram et al. 2006; Bender and Ovtcharova 2021; Bender, Trat et al. 2022; T. Chen 2009; Gyulai, Pfeiffer, Nick et al. 2018; Kramer, C. Wagner et al. 2020; Liang 2010; Lim, Yusof et al. 2019; Lingitz, Gallina et al. 2018; Meidan, Lerner, Hassoun et al. 2009; Meidan, Lerner, Rabinowitz et al. 2011; Mori and Mahalec 2015; Öztürk, Kayalıgil et al. 2006; R. Zhang and Wu 2012;Cos Juez, Nieto et al. 2010; Lu, Ramaswamy et al. 1994; Ren, Chaw et al. 2022; Schneckenreither, Haeussler et al. 2021; Yamashiro and Nonaka 2021*
Production analysis	Analysing the production performance based on historical lead time data	Lead time data can be used to identify patterns and trends in the production that potentially could be improved. By utilizing LTP, the company may be able to take proactive actions that can improve the production efficiency.	*Ahmarofi, Ramli et al. 2017; Cao and Ji 2021; Chien, C.-Y. Hsu et al. 2012; Choueiri, Sato et al. 2020; G. M. Lee and Gao 2021; Meidan, Lerner, Hassoun et al. 2009; Meidan, Lerner, Rabinowitz et al. 2011; Pfeiffer, Gyulai et al. 2016; J. Wang, J. Zhang and X. Wang 2018*
Production monitoring	Monitoring and controlling the production based on lead time data	LTP could be employed to predict remaining lead times at the shop floor. The result could further be utilized to update the customers on order status. Remaining lead times could also be used to make appropriate adjustments to job priorities on the shop floor, in order to ensure that all orders are delivered on time.	*T. Chen 2009; Gyulai, Pfeiffer, Bergmann et al. 2018; Gyulai, Pfeiffer, Nick et al. 2018; S. Huang, Guo et al. 2019; Meidan, Lerner, Hassoun et al. 2009*

Table 3: Summary of the identified application areas for LTP.

3.2.3 Prior methods used for lead time prediction

There are various methods used for LTP. Some of the methods mentioned in the literature are rule-based approaches and rely on data manipulation rules or policies created by humans, or they can be based on quantitative techniques such as simulation and queuing theory (Gyulai, Pfeiffer, Bergmann et al. 2018). According to Alenezi, Moses et al. (2008), some policies evaluate lead time as a linear function of a single order attribute, which is not very efficient given how closely related lead times are to system parameters and system status. Many of these traditional approaches are used to determine the average lead time based on historical data without taking lead time variability into consideration (Alenezi, Moses et al. 2008; T. Chen 2009; Lingitz, Gallina et al. 2018). This may result in inaccurate estimations because lead times in real-life frequently vary greatly. One of the most common ways to predict lead times for manufactures today is still based on the production managers prior expertise, and information gathered from the fabric operators (Mourtzis, Doukas et al. 2014; Schneckenreither, Haeussler et al. 2021).

S. Huang, Guo et al. (2019) states that current MTO manufacturing systems must be designed with quick response, wise decision-making and exact execution in mind in order to keep up with today's competitive market. Machine learning has been highlighted as an emerging technology that is able to handle the uncertainty and variability in today's production systems. According to Mourtzis, Doukas et al. (2014), ML is one of the most reliable techniques for predicting lead times, due to its ability to examine an extensive amount of complex data containing several parameters and non-linear connections. This thesis will only focus on ML methods used for LTP, and examples of various ML methods applied for this task are examined in section 3.4.

3.3 Machine learning

ML is a branch of artificial intelligence (AI). The fundamental idea of ML is to enable computers to learn without being explicitly programmed (Géron 2022). Data can be transformed into feature spaces that can be utilized for prediction, detection, classification, regression, or forecasting by using ML algorithms to identify highly complex and non-linear patterns in the data (Wuest, Weimer et al. 2016). The performance of a ML model depends heavily on the quality of the data, and Jain, Patel et al. (2020) emphasizes how crucial it is to comprehend and improve the data for the model to produce outcomes that are accurate and dependable. In addition to high-quality data, a sufficient amount of data is also required in order to properly train the algorithms to generate more precise outputs. However, this amount is dependent on the type of ML model (Grolinger, Hayes et al. 2014).

The algorithms learn different data patterns through training. The training process refers to how they automatically adjust their behavior through repetition in order to get better and better at doing the desired task. The objective is to have the model deliver the intended result based on unobserved data (El Naqa and M. J. Murphy 2015). It has been demonstrated to be an approach with great potential for managing complex systems, and various academics point to ML as a critical component for present and future PPC challenges (Gyulai, Pfeiffer, Bergmann et al. 2018; R. Murphy, Newell et al. 2019).

3.3.1 Types of Machine learning

ML models can be classified based on how they are supervised during training (Géron 2022). The three main learning approaches for ML are supervised learning, unsupervised learning and reinforcement learning, see Figure 4. The degree of supervision the ML model receives during training is the basis for these definitions. Supervised learning involves training a model with labeled data, where input-output pairs are provided, enabling the model to learn patterns and make predictions or classifications. Unsupervised learning, on the other hand, deals with unlabeled data, where the model seeks to identify underlying patterns or structures without explicit guidance. Reinforcement learning focuses on an agent learning to make sequential decisions by interacting with an environment and receive rewards or penalties based on its actions. The obtained data is used to learn a policy or strategy that maximizes the reward. Supervised learning is the focus of this paper.

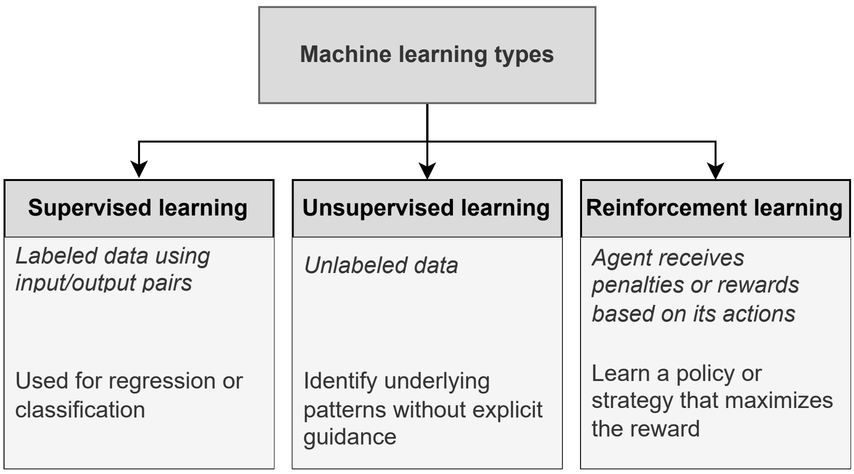


Figure 4: The three main learning approaches for machine learning

Supervised learning

As mentioned above, in supervised learning the model is trained using labeled data, meaning that the target values are included in the dataset. Supervised machine learning models are able to identify dependencies and patterns between the input data and the target value. Typical use cases for supervised learning are classification and regression (Géron 2022).

Classification

In a classification problem the training dataset contains input/output pairs, where the output are discrete class labels. After training, the model should be able to learn the classes based on the provided input dataset. One example of a classification problem in the manufacturing industry is quality control in product inspection. The objective in this scenario is to categorize products into several quality groups according to their traits and features. For example, the system can collect images or sensor data of manufactured items and train a classification model to categorize them as "acceptable" or "defective."

Regression

In a regression problem the training dataset includes input/output pairs with continuous values as the output. The ML model is used to detect patterns between the input and the output, and make predictions for new input data without the corresponding target value based on the detected relationship (Maharana, Mondal et al. 2022). In the manufacturing industry, LTP could be an example of a regression problem. By providing the ML inputs consisting of lead time affecting features, e.g. component length and WIP level, along with the corresponding lead time, the model should be able to predict the lead times of new orders based on the detected patterns from the training.

3.4 Machine learning for lead time prediction

Machine learning is an explored field for LTP. R. Murphy, Newell et al. (2019) refers to ML as a technique with great potential for managing complex dynamic systems. ML can enhance the performance of manufacturing tasks, such as LTP, according to previous studies. Gyulai, Pfeiffer, Bergmann et al. 2018 argues that the use of AI in manufacture will be a crucial component for future PPC strategies. Schkarin and Dobhan (2022) carried out an interview study where they interviewed nine experts with knowledge within the topics of AI and production planning. The experts all agreed that AI is gaining ground as a method used for production planning, and 80% of them implied that it can help increase the efficiency of processes within the production planning.

The application of ML for LTP is given a special focus in customer specific environments, which are arising due to the increased demand of customized products (Schuh, Gützlaff, Sauermann, Kaul et al. 2020). As mentioned, the majority of the conventional lead time estimation techniques use static approaches that calculate the average lead time from historical data, without incorporating the highly scattered lead time that arises in MTO environment (Lingitz, Gallina et al. 2018). The large product variety in these environments increase product and production complexity and uncertainty. ML is a technique specially suited to handle high complexity data with non-linear relations among the variables, which are common characteristics for MTO companies (Asadzadeh, Azadeh et al. 2011).

Previous research have explored a wide variety of ML models for LTP. The models are tested and evaluated on different case companies, and a generalized model selection applicable to all companies does not exist (Kramer, C. Wagner et al. 2020). The most appropriate model relies on the process being studied and the parameter that has to be predicted (Gyulai, Pfeiffer, Nick et al. 2018). Therefore, Kramer, C. Wagner et al. (2020) argued that it is not sufficient to investigate only one model, but to compare several models and find the best suited for the manufacturing environment under study. R. Murphy, Newell et al. (2019) proved the same in their study. Kramer, C. Wagner et al. (2020) summarized the current ML models used for LTP as linear regression (LR), decision tree (DT), random forest (RF), k-nearest neighbor (k-NN), support vector regression (SVR), lasso regression (LassoReg), ridge regression (RidgeReg), artificial neural network (ANN), multivariate adaptive regression (MARS), deep neural networks (DNN), bagging and boosting decision trees (BDT).

Previous literature has proven that ML techniques are superior to conventional methods. Gyulai, Pfeiffer, Nick et al. (2018) compared LR, RF, DT and SVR for LTP in a flow-shop manufacturing optical lens. In this case, RF achieved the highest accuracy. Additionally, the ML models were compared to analytical approaches and proved to outperform the conventional methods. Asadzadeh, Azadeh et al. (2011) compared ANN, fuzzy regression and conventional regression for LTP, and concluded that ANN is superior to the other methods. The authors added that fuzzy- and conventional regression methods are not able to handle the complex non-linear data.

The appropriate ML model is dependent on the manufacturing system. In the study of Kramer, C. Wagner et al. (2020) the authors compared LR, DR, DT, RF, BDT and DNN for LTP in a job shop. The authors stated that RF often provide best results in the literature, however in their study DNN achieved the best performance, which exemplifies that the appropriate model is context-specific. Flowtime estimation for different manufacturing systems were explored in the study of R. Murphy, Newell et al. (2019). ANN, RF, BDT and rule-based Cubist method were evaluated in a simulated flow shop and hybrid flow shop with different flow policies and dispatching rules. The best performing model varied depending on the simulated environment. In general, the results demonstrated that factors such as flow of the manufacturing line, processing times at machines and the order characteristics all have an impact on the suitable ML model. RF, ANN and Cubist all proved to achieve high results in different contexts.

RF has proved to be a suitable ML model for LTP in several studies. In the study of Schuh, Gützlaff, Sauermann, Kaul et al. (2020), the authors explored DT, k-NN, ridge regression and RF for the prediction of order-specific transition times, and RF proved to be the best model. Lingitz, Gallina et al. (2018) compared eleven different ML models for LTP in a hybrid MTS/MTO job shop environment withing the semiconductor industry. Once again, RF provided the best result. In the research of Fussenegger and Lange (2022), methods for LTP were investigated in the scope of explainable AI. RF, DT, LR, ridge regression, lasso regression, elastic net and SVR were examined. In their research they concluded that LR and SVR were not suitable due to their inability of handling non-linear relationship between the variables. The best performing model was again RF. However, the authors highlighted that DT provided highest interpretability, followed by RF. RF also achieved the highest accuracy in a simulated MTO flow shop environment, compared to DT and LR in a study conducted by Pfeiffer, Gyulai et al. (2016). The production time of a textile factory was investigated in Atik, Kut et al. (2021). In their study a comparison between SVR, RF, DT and bagging decision tree was performed. The bagging decision tree and RF showed the highest predictive performance.

ANN is also a frequently applied ML model for LTP. Mezzogori, Romagnoli et al. (2019) explored multivariate LR and ANN to estimate the lead time of incoming jobs in an MTO workload control (WLC) system. The WLC system based on ANN forecasting achieved the best results. Gacek (2018) also compared multivariate LR to ANN in order to predict the total production flow time in an MTO environment. Ahmarofi, Ramli et al. (2017) predicted the completion time for a production line using both ANN and LR. In both aforementioned studies, ANN outperformed LR. Chien, C.-Y. Hsu et al. (2012) predicted the cycle time in a semiconductor industry with a model combining gauss-newton regression method and ANN. An inclusion of ANN resulted in improved accuracy compared to a regular gauss-newton regression method. In the study of Meidan, Lerner, Hassoun et al. (2009) a Naive Bayes classifier (NBC) was developed for cycle time prediction. As stated in the article, NBC was chosen due to it's simplicity, even though ANN is more accurate and DT is more intuitive. In a later study by the same authors they compared NBC to ANN, DT and multinomial logistic regression (MLR) (Meidan, Lerner, Rabinowitz et al. 2011). ANN and MLR proved to achieve the highest accuracy. Cao and Ji (2021) investigated the production cycle time of a garment industry, where it is hard to obtain precise cycle time due to the large amount of manual labour. The cycle time was successfully predicted with ANN. Susanto, Tanaya et al. (2012) explored the standard lead time of a textile factory using ANN. As described in the study, textile factories are also influenced by poor data records of the production time, due to the extensive amount of manual labour. The authors concluded that results from the ANN model could successfully be employed in the case company. The results proves the strength of ANN's ability to handle complex data structures and capture underlying relationships among variables (Ahmarofi, Ramli et al. 2017).

Even though RF and ANN has proved to be superior over several ML models, several authors have also achieved high accuracy with other ML models as well. Alenezi, Moses et al. (2008) developed a SVR for real-time flowtime prediction in a multi-resource, multi-product system. The method is compared to time-series models and ANN. The results showed that SVR is most suited for large-scale production, and ANN and SVR results are rather similar for the small and medium production system. SVR was also successfully applied in a case study of the aerospace industry for predicting if the lead time for a batch of jobs could be finished within a given time horizon in the study of Cos Juez, Nieto et al. (2010). Backus, Janakiram et al. (2006) explored cycle time prediction for both the whole production and intermediate production steps with clustering, k-NN and DT. The models were compared in terms of feasibility to maintain and rebuild the model, where DT achieved the best results. Bender, Trat et al. (2022) investigated the use of automated machine learning (AutoML) for LTP, which is a way of automating the processes in a ML-life cycle, from data cleaning to deployment. They highlight that AutoML is capable at algorithm selection, training, hyperparameter optimization and benchmarking, but more effective methods are still required to achieve data understanding, transformation, filtering, pre-processing and feature engineering. Additionally, the authors concluded that AutoML is best suited if the input data is clean and of high-enough quality. In a future research, Bender and Ovtcharova (2021), further explored the approach of AutoML for LTP by integrating it into an existing enterprise system. The approach was technically feasible and beneficial for order scheduling, but more focus is needed towards better integration of automated model monitoring and retraining.

It is important to investigate factors affecting lead time in order to define suitable input features for the ML model. Meidan, Lerner, Rabinowitz et al. (2011) summarized the most common lead time factors identified in the literature as: work-in-process (WIP) levels, line-bottlenecks, dispatching rules, lot priority and scheduling policies, equipment load, and product mix. J. Wang and J. Zhang (2016) used a dataset with order characteristics, i.e. processing times and priority, and workshop conditions, i.e. waiting queue length, utilization of machine and WIP level, for LTP. In the research of S. Huang, Guo et al. (2019), the authors investigated a case company equipped with RFID scanners for inn-out pairs for all machines and for the buffer areas. The RFID system provided them great access to retrieve the system status information, such as processing time at each machine, waiting time in buffers, queuing length and type, and transfer time. In the work of Asadzadeh, Azadeh et al. (2011) they predicted the lead time using three influential factors: sum of failures, sum of repairs and sum of processing time for all operations, and concluded that failures due to machine breakdowns were the most crucial parameter. Li, Yang et al. (2015) conducted a simulation experiment to investigate the relative importance of predictor variables for lead time. The results showed that WIP variables, machine status (busy or idle, up or down and processing time) of important workstation, customer order size and forecasted arrival rate of future orders were the most influential factors for the ML model, and from these the WIP variables and machine failures were the factors that had the most significant impact.

A summary of the explored ML models in the reviewed literature is presented in Table 4. Models that only occurred in a single article are excluded from the table. The best performing model is presented with a checkmark, or in bold writing if it is categorized under “other”. The remaining models that were explored are marked with an “X”, the same applies for articles that has not concluded with a best performing model or only investigated one model. The “other” category contains models that has not occurred several times in the reviewed articles. Additionally, the table is categorized as followed: DNN incorporate deep autoencoder and deep belief network. DT incorporate bagging and boosting decision tree. The hybrid neural network, fuzzy c-means-back propagation network is categorized as a subfield of ANN.

Source	ANN	RF	LR	DT	SVR	DNN	RidgeReg	LassoReg	k-NN	NBC	Cubist	Other
Ahmarofi, Ramli et al. 2017	✓		X									
Alenezi, Moses et al. 2008	X				✓							
Asadzadeh, Azadeh et al. 2011	✓		X									FuzzyReg
Atik, Kut et al. 2021		✓		✓	X							
Backus, Janakiram et al. 2006				✓					X			Clustering
Bender and Ovtcharova 2021												AutoML
Bender, Trat et al. 2022												AutoML
Cao and Ji 2021	X											
T. Chen 2009	X											
Chien, C.-Y. Hsu et al. 2012	X											
Cos Juez, Nieto et al. 2010					X							
Fang, Guo et al. 2020						X						
Fussenegger and Lange 2022		✓	X	X	X		X	X				
Gacek 2018	✓		X									
Gyulai, Pfeiffer, Nick et al. 2018		✓	X	X	X							
Gyulai, Pfeiffer, Bergmann et al. 2018		X										
S. Hsu and Sha 2004	X											
S. Huang, Guo et al. 2019						X						
Kramer, C. Wagner et al. 2020		X	X	X		✓						
G. M. Lee and Gao 2021	X											
Lingitz, Gallina et al. 2018	X	✓	X	X	X		X	X	X			MARS

Source	ANN	RF	LR	DT	SVR	DNN	RidgeReg	LassoReg	k-NN	NBC	Cubist	Other
Meidan, Lerner, Hassoun et al. 2009										X		
Meidan, Lerner, Rabinowitz et al. 2011	✓			X						X		**MLR**
Mezzogori, Romagnoli et al. 2019	✓		X									
R. Murphy, Newell et al. 2019	X	X		X							X	
Öztürk, Kayalıgil et al. 2006			X								X	
Pfeiffer, Gyulai et al. 2016		✓	X	X								
Schneckenreither, Haeussler et al. 2021	X											
Schuh, Gützlaff, Sauermann, Kaul et al. 2020	X	✓		X			X		X	X		
Susanto, Tanaya et al. 2012	X											
J. Wang and J. Zhang 2016	X											
J. Wang, J. Zhang and X. Wang 2018	X											

Table 4: Summary of explored machine learning models for lead time prediction from the literature review.

Previous studies has shown that ML models is an emerging research field with a lot of potential for LTP. It is a useful technique for predictive analysis due to its capacity to analyze a large amount of complex data involving several parameters with non-linear relationships. ML can increase the efficiency of a production system and help encapsulate the dynamics of the system. Previous research has shown that ML models are capable of outperforming conventional techniques, providing higher accuracy and lower error. A generalized ML model for LTP does not exist and the suitable model is context dependent, therefore multiple models should be tested. RF and ANN has showed superior performance compared to other ML models in a number of studies, these are therefore chosen to be further explored in this study.

3.4.1 Artificial neural networks

Artificial neural networks are inspired by the structure and functioning of biological neural networks in the human brain (Krogh 2008; Russell 2010). By replicating the operations of biological neurons, ANNs are able to solve many sorts of problems such as classification, prediction and speech recognition (Krogh 2008).

The artificial network is composed of differently connected neurons. The neurons are connected into input layers, hidden layers, and output layers, which together form the artificial network or architecture of an ANN, presented in Figure 5. The input layer receives input data, which is then passed through the hidden layers, where complex computations occur. The output layer provides the final output of the network, which could be a classification, regression, or any other desired prediction.

An ANN's architecture can also be divided into recurrent and feed-forward networks. The two of them differ from one another in terms of how they connect to other neurons. Recurrent networks feature at least one feedback loop in which the output of one neuron is used as input for the same neuron, whereas feed-forward networks only have connections going in one direction (Russell 2010). The focus of this thesis is feed-forward network, and a simplified model of it's architecture is shown in Figure 5

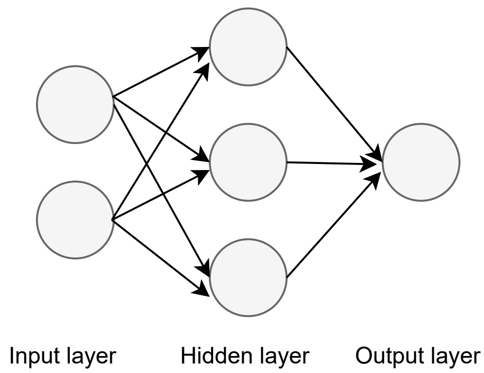


Figure 5: Architecture of a feed-forward artificial neural network

The neurons are mathematical functions in an ANN, presented in Figure 6. A neuron receives several inputs, and calculates the weighted average, and sends this weight through a nonlinear function, referred to as the activation function. Each neuron have specified weights and thresholds, the activation function determine whether or not the neuron's output exceeds the threshold. When a neuron's output exceeds the threshold, it is triggered and sends data to other neurons in the next layer of the network (Krogh 2008). The weights are generated randomly and modified during the model's training phase (Krogh 2008). Therefore the main purpose of training is to find optimal weights for all neurons, and greater weights are given to more significant inputs than those deemed less necessary.

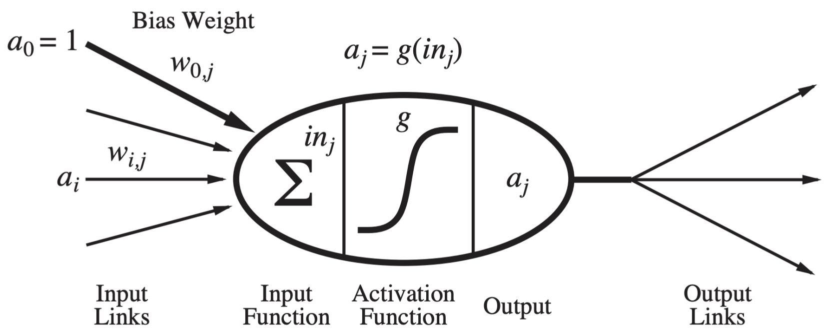


Figure 6: A model of a neuron derived from Russell (2010)

The adjustment of neurons weights are typically done using a technique called back propagation, which is a technique that compares the output of the network with the desired output and updates the weights accordingly until the variance between actual and desired output are minimized (Ahmarofi, Ramli et al. 2017; Krogh 2008). The training process enables ANNs to gradually enhance their performance for a specific tasks (Krogh 2008). The most common ANN architecture for LTP is a multilayer perceptron feed-forward network with back-propagation learning algorithm, from here referred to as a multilayer perceptron (MLP) model (Ahmarofi, Ramli et al. 2017; Cao and Ji 2021; Gacek 2018; Susanto, Tanaya et al. 2012). The MLP models have several hidden layers between the input and output layer, and are able to detect intricate patterns in the data.

3.4.2 Random forest regressor

Random forest is a popular ML model that can be used for both classification and regression (Rodriguez-Galiano, Sanchez-Castillo et al. 2015). This thesis will focus on RF regression which is an extension of decision tree regression (Atik, Kut et al. 2021).

DTs are hierarchical models with a tree-like structure made up of nodes and branches, see Figure 7. DTs recursively split the data based on feature values to make predictions in a hierarchical and interpretable manner. The internal nodes of the tree represent tests or rules that is performed on input values and the branches represent the outcome of these tests (Russell 2010). The data splitting process is continued from the root node until a previously determined stop condition is met (Rodriguez-Galiano, Sanchez-Castillo et al. 2015). The leaf nodes of the tree have a simple regression model attached to it, which is used to predict a value that will be returned by the DT. A pruning process is applied in order to reduce the complexity and improve the generalization capability of the tree (Rodriguez-Galiano, Sanchez-Castillo et al. 2015).

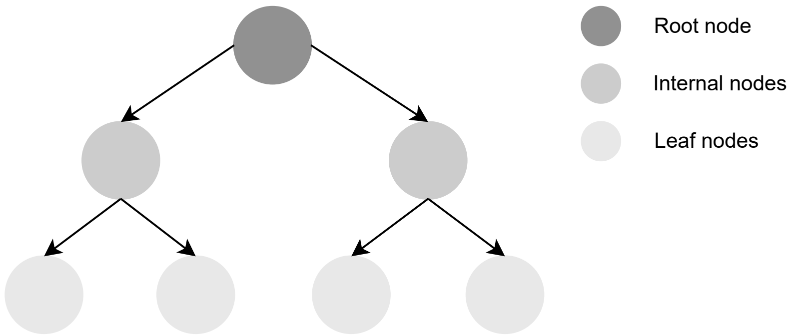


Figure 7: Architecture of a decision tree.

RF is an ensemble learning method that combine several randomly generated DTs which are built independently without interaction, and run in parallel (Atik, Kut et al. 2021). The algorithm combines the predictions of multiple DTs to make more accurate and robust predictions (Rodriguez-Galiano, Sanchez-Castillo et al. 2015). When given an input x, the RF model constructs a set of K different regression trees before averaging the sum of the outputs from the various decision trees (Rodriguez-Galiano, Sanchez-Castillo et al. 2015). The mathematical expression is shown below:

$$\hat{f}^K(x) = \frac{1}{K}\sum_{k=1}^{K} T_k(x)$$

All the distinct trees are trained on different subset of the training data in order to prevent correlation between the trees (Atik, Kut et al. 2021; Rodriguez-Galiano, Sanchez-Castillo et al. 2015). This process is called bagging. In the bagging process, new training sets are generated using randomly selected data points from the original dataset with replacement. This means that some instances may appear multiple times in a training sample, while others may be left out (Rodriguez-Galiano, Sanchez-Castillo et al. 2015). The bagging process creates diversity in the training data for each decision tree.

3.5 Data quality assessment

The application of ML in manufacturing environments is highly data dependent. Through the use of auto-identification and telemetry data collection systems, such as RFID, several manufacturing industries are able to produce the amount of needed data (Oluyisola 2021). These technologies are becoming more ubiquitous for manufacturing firms, due to reduced cost and availability (Oluyisola 2021, derived from Iansiti and Lakhani 2014).

Various automated data collection systems are available on the market today. Perez (2014) has looked into several systems for collecting location data in manufactures. *Bar code systems* locate a specific part type or operation using manual scanners and bar codes (Cox and Blackstone 2002). The method's primary flaw is that it depends on the operators to do the scanning, which increases the likelihood of human mistakes (Bender, Trat et al. 2022; Perez 2014). Computer vision systems use cameras to record video data across the factory. Although this can function without human interaction, it requires a high level of data science expertise to function effectively (Perez 2014). RFID is a method that uses electronic tags to make automated component tracking and routing through the manufacturing process (Cox and Blackstone 2002; Sabbaghi and Vaidyanathan 2008). The data is captured through readers and is transmitted through the air. This is an easy system to install, maintain and run autonomously. The main drawback is that it is sensitive to noise in metallic environments (Perez 2014).

These sensor systems are one of several foundations in the digitalization of manufactures. Sjödin, Parida et al. (2018) conducted a review to address the needs before enabling intelligent manufacturing technologies. The review stressed the need of accurate data collection, in order to ensure that the right data is acquired and reduce signal interference, it was further claimed that improving data creation and data quality was essential. High data quality is becoming an increasingly important topic of study for practitioners and researchers as interest in using intelligent technology grows (Cichy and Rass 2019; Lindström, Persson et al. 2023; Sjödin, Parida et al. 2018). The success of applying intelligent concepts, such as ML, depends heavily on the quality of the utilized data.

A general definition of data quality is difficult to obtain because of the variety of data, its use cases, and its dimensions (Lindström, Persson et al. 2023). A common definition for high data quality is that the data must satisfy the users' requirements (Strong, Y. W. Lee et al. 1997). Several data quality frameworks has been developed in the literature to evaluate and enhance data quality, due to the significance of achieving high-quality data (Cichy and Rass 2019). A general framework for data quality incorporate the steps of data quality definition, data quality assessment and data quality improvement (Cichy and Rass 2019). In the context of ML and big data (BD), models like the proposed CRISP-DM are offered to improve the data quality through data preparation, however these frameworks do not explain how to address the key issues with data quality (D. C. Corrales, J. C. Corrales et al. 2018; D. C. Corrales, Ledezma et al. 2018).

It is common in the literature to describe data quality through data dimensions which can reflect the overall quality of the data (Cichy and Rass 2019; Lindström, Persson et al. 2023). Cichy and Rass (2019) has conducted a literature review of data quality frameworks. The authors concluded that the most common dimensions in the literature are: *Completeness:* Determining whether there exists sufficient and suitable amount of data for the task at hand. *Accuracy:* The evaluation of whether the data is correct, reliable and certified. *Timeliness:* Assessment of the datas' age and if it is appropriate. *Consistency:* The extent to which the data is consistent with earlier data. *Accessibility:* Determining whether the data is readily, rapidly, and easily retrievable. Despite the fact that the dimensions help define data quality, it might be challenging to specifically specify a dimension for the needed use case or tie it to a company's application (Perez 2014). Juddoo (2015) conducted a research with the focus of mapping data quality frameworks specifically to BD. They concluded that the most relevant data dimensions for BD were *consistency*, *completeness*, *accuracy* and *believability*. The latter dimension specifies whether a data source generates true, reliable and credible data (Juddoo 2015). The authors also emphasize that the most crucial dimension is context specific, exemplifying that for sensors and RFID the accuracy and completeness dimensions are likely to be the most important. A summary of the data quality dimensions are presented in Table 5, along with the occurrence of each dimension in the reviewed articles by Cichy and Rass (2019).

Data dimension	Definition	Occurance	Source
Completeness	Determining whether there exists sufficient and suitable amount of data for the task at hand	10	Cichy and Rass 2019; Laranjeiro, Soydemir et al. 2015; Taleb, El Kassabi et al. 2016
Accuracy	The evaluation of whether the data is correct, reliable and certified (close to real-world events and error-free)	8	Cichy and Rass 2019; Laranjeiro, Soydemir et al. 2015; Taleb, El Kassabi et al. 2016
Timeliness	Assessment of the datas' age and if it is appropriate	9	Cichy and Rass 2019
Consistency	The extent to which the data is consistent with earlier data and data constraints	5	Cichy and Rass 2019; Taleb, El Kassabi et al. 2016
Accessibility	Determining whether the data is readily, rapidly, and easily retrievable	5	Cichy and Rass 2019
Believability	Specifies whether a data source generates true, reliable and credible data	3	Juddoo 2015; Laranjeiro, Soydemir et al. 2015

Table 5: Data quality dimensions

A data quality assessment determines how well each dimension is performing (Cichy and Rass 2019). The assessment involves selecting and developing data quality measurement types that are subsequently used on the available data. The measurement types are usually subjective evaluation and/or quantitative metrics. Subjective assessment is achieved through communication or questionnaires with the data consumer, where they rate the degree of quality of the dimension (Cichy and Rass 2019). Metrics frequently measures the number of percentages of some constraints. Most metrics has an outcome between 0 and 1, with 0 denoting the improper value and 1 denoting the appropriate value Juddoo 2015; Pipino, Y. W. Lee et al. 2002. Metrics are task-independent or task-dependent, depending on whether the contextual factor is considered (Pipino, Y. W. Lee et al. 2002). Three functional forms for data quality metrics are presented in the literature for data quality assessment, presented in Table 6 (Pipino, Y. W. Lee et al. 2002). Utilizing these functional forms, one can derive metrics suitable for a particular use case. The metrics are also well suited for BD, however a constant stream of data may present difficulties due to the challenge of measuring the number of outcomes (Juddoo 2015).

Functional form	Description	Dimension
Simple Ratio	Simple ratio is the relation between desired and undesirable outcomes, where 1 represent the most desirable, and 0 the least. For a numerical attribute a range of acceptable values can be determined, and the undesirable outcomes will be the variables that does not fall under this umbrella.	Completeness, Consistency, Accuracy
Min or Max operation	Min or Max operation is used to deal with various data quality indicators (variables). The formula calculates the minimum or maximum value among the normalized variables. The min operator assigns the dimension the value of the weakest data quality indicator, and the max operator the strongest.	Min operator: Believability, Max operator: Timeliness, Accessibility
Weighted average	Weighted average is applied for multivariate cases. The process allocates weighting factors to signify the variables' significance in the assessment of a dimension. The company needs a good understanding of each variable for the overall evaluation of the dimension.	Believability

Table 6: The functional forms and coherent dimensions, presented by Pipino, Y. W. Lee et al. (2002)

Taleb, El Kassabi et al. (2016) presents the simple ratio (SR) for the data quality dimensions completeness, consistency and accuracy for structured schema data which they have rewritten as followed:

$$Comp = \frac{Nr.\ of\ missing\ values}{total\ nr.\ of\ values\ in\ the\ dataset}$$

$$Cons = \frac{Nr\ of\ values\ that\ respect\ the\ constratint}{total\ nr.\ of\ values\ in\ the\ dataset}$$

$$Acc = \frac{Nr.\ of\ correct\ values}{total\ nr.\ of\ values\ in\ the\ dataset}$$

The major objective is to improve the quality of the data after conducting an evaluation of it. A root cause analysis is frequently employed in conjunction with the assessment of data dimensions (Cichy and Rass 2019). In a root cause analysis the source to poor-data is further examined. Strong, Y. W. Lee et al. (1997) defines poor-data quality as any negative affects on one or more of the dimensions, which render the data fully or partially unusable. The identified issues might incorporate both process- and data-driven improvements (Cichy and Rass 2019). Data-driven improvement is achieved by directly modifying the data, while process-driven improvement concerns process control, where the data is checked and managed along the manufacturing process. Process redesign is also a part of process-driven improvement, where a redesign can help eliminate the cause to low data quality (Glowalla, Balazy et al. 2014). Data-driven improvement in the scope of ML and BD is frequently called *data cleaning* or *data preparation* (D. C. Corrales, Ledezma et al. 2018; Gudivada, Apon et al. 2017; Juddoo 2015).

In the literature review by D. C. Corrales, J. C. Corrales et al. (2018) they identified five data quality issues, in the scope of ML regression tasks. These include missing values, outliers, high dimensionality, redundancy, and noise. Noise is a general term for irrelevant and meaningless data, which can incorporate the other issues. In the article of Taleb, El Kassabi et al. (2016), BD quality issues were mapped to the corresponding data quality dimension, derived from Laranjeiro, Soydemir et al. (2015) and M. Chen, Song et al. (2012). The detected issues were highlighted as missing data, incorrect data, data entry errors, irrelevant data, outdated data, misfiled and contradictory values, uniqueness constraints and functional dependency violation, lack of integrity constraints, wrong data type and poor schema design.

Alwan, Ciupala et al. (2022) conducted a systematic review on data quality challenges related to sensor systems, with the focus on cyber-physical systems (CPS) which refers to interconnected networks that integrate the physical and virtual world. The detected quality issues were sensor nodes malfunctions, calibration issues, poor sensor nodes quality, environmental effects, external noise, networks or communication errors, and real-time scheduling problems. The literature classified these issues into three main categories: i) Errors in sensor nodes measurement due to the devices calibration, accuracy and adaptive sampling, ii) hardware failures in sensor nodes or communication networks, and iii) mismatches in sensor nodes spatial and time parameters. Another literature review of data quality issues in the scope of internet of things (IoT), which refers to the network of interconnected devices and objects that can communicate and exchange data with each other over the internet, was conducted by Mansouri, Sadeghi Moghadam et al. (2021). They detected the following issues; duplicate, noisy data, errors, lack of context, outliers, response time, inconsistency, redundancy, uncertainty, ambiguity, insecurity, source ambiguity, and incompleteness. Error refers to any faults that occur during the data stream's capture or processing owing to sensor failures brought on by unusual environments, aging, network data exploitation, improper installation, or usage.

In the study of Lindström, Persson et al. (2023) data quality issues are detected in the scope of PPC. Eight respondents contributed to the findings from various sectors within mechanical and electrical engineering to the oil and gas industry. The detected data quality problems were inaccurate data entries, changing task needs, data production errors, multiple sources of data, distributed system, and lack of computing resources. Multiple causes were highlighted, which are presented in Figure 8. Lindström, Persson et al. (2023) has summarized causes for poor-data quality in organizations, derived from Eckerson (2002) and Eppler (2006) as; outdated or obsolete data, manipulation of stored data, spelling or typing errors, incorrect data entries, duplicates or multiple data sources, mismatched syntax or formats, wrong data coding and tagging, lack of referential integrity checks, and poor system design.

Data quality issues are detected in several domains. Figure 8 summarizes the main data quality issues described in the aforementioned literature, and their coherent data quality dimension where that is specified. The figure separates problems regarding ML and BD, CPS and IoT, and PPC and organizations.

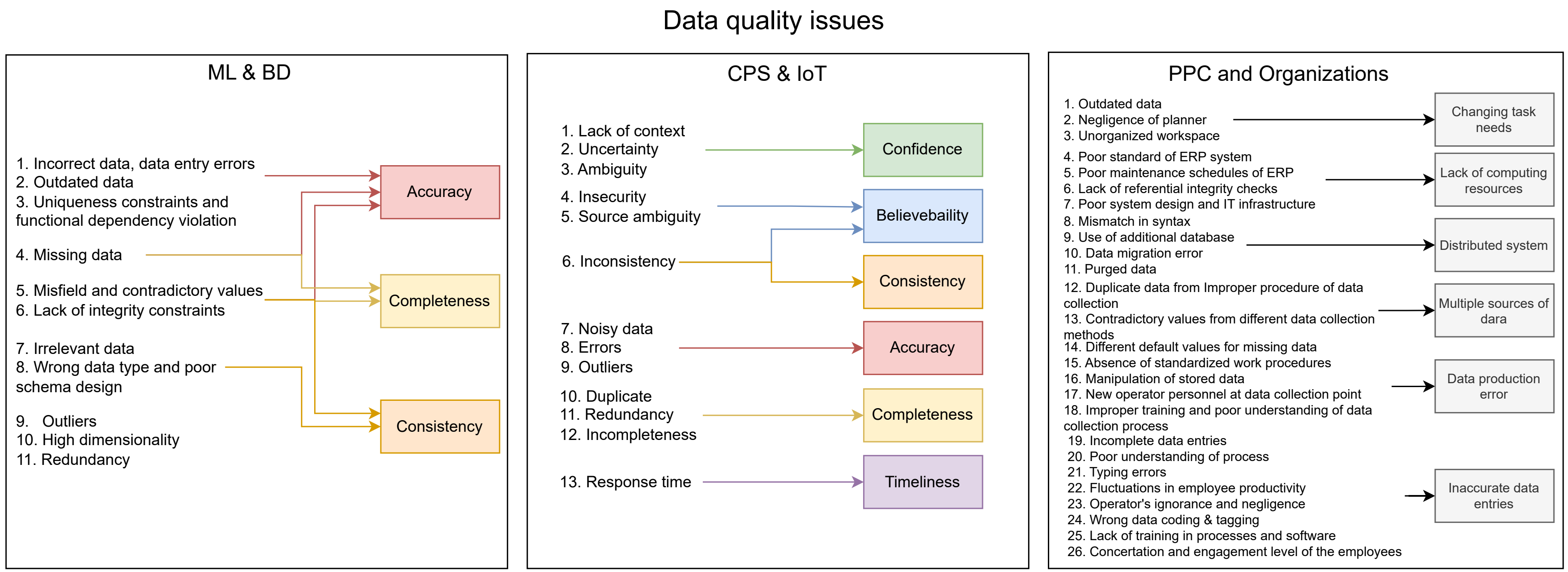


Figure 8: Data quality issues and it's belonging dimension, where it is specified, from 1) ML & BD: D. C. Corrales, J. C. Corrales et al. (2018), Laranjeiro, Soydemir et al. (2015) and Taleb, El Kassabi et al. (2016), 2) CPS & IOT: Alwan, Ciupala et al. (2022) and Mansouri, Sadeghi Moghadam et al. (2021), 3) PPC & Organizations: Lindström, Persson et al. (2023)

4 Case study

This section introduces the case company for the thesis. The case study is a continuation of previous work conducted by the authors, as mentioned in Section 2. It is conducted in order to achieve business understanding, which is the first phase of the ML process. The case company, Gilje, is introduced in section 2.2. The case study provides a more detailed description of the company, and aims to further identify the company's characteristics and map their existing planning approach. Data provided from the company is examined and used for machine learning based lead time prediction.

4.1 Market

The window industry is characterized by high demand uncertainty. Windows are becoming more and more tailor with individual design needs. This increases the importance of a flexible production processes. Gilje has noticed an increased competition from Eastern Europe and other low-cost country, where the production costs are lower. This increases the importance of focusing on high quality, innovative and sustainable products. In order to fulfill this, Gilje is dependent on efficient PPC, with quick response and high delivery precision.

4.2 Production planning

The production planning process begins when a consumer places an order through Gilje's online store or by email with a desirable delivery date. The production planner views the order and checks the capacity occupation of the given order, and compare it to the available capacity that day. If the due date is feasible, a price offer is sent back to the customer. If the due date is not feasible, a new due date is proposed to the customer. When both parties agree on the price and due date, the order is accepted and scheduled for production.

The capacity occupation of a window is determined by the *glass points* of a window. The amount of glass points a window receives depend on the number of glass panes, the size of glass panes, special characteristics and the planners past experience. The capacity status for the production is visualized through a bar chart, see Figure 9. The bar chart contains the glass points of all windows scheduled to be produced a specific day. When the bar chart reaches 100% the maximum capacity for a day is reached. To assess the availability for a given day the total amount of glass points can not exceed the maximum capacity. Maximum amount of glass points for a day is 620 glass points. This is a floating number over a period of time. Changes in the maximum capacity level is determined by experience and production planners observations from the production.

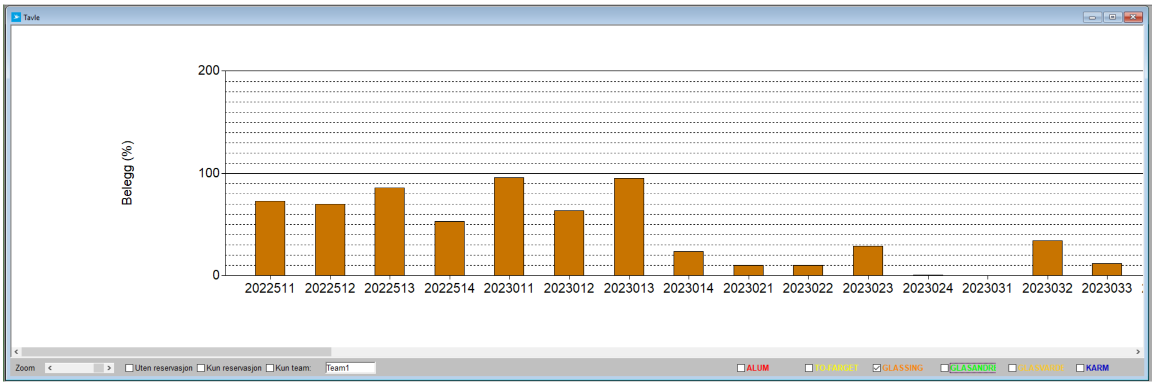


Figure 9: The bar chart from Gilje's production planning. Each bar represent the total glass point occupation for a day.

The orders are placed in the bar chart according to their due date. The due date is typically a few weeks or months after the receipt of the order. To determine the start day of production Gilje uses backward scheduling from the due date. Based on past experience Gilje has noted that the lead time of an order is between one and two days, and can even be as low as six hours. For scheduling Gilje assume one to two days of production and thereafter include five buffer days at the finished goods inventory. The buffer days are included in case the production is delayed or interruptions appear. Therefore the production of an order starts around seven days before the due date. The total time from production start to termination, namely the lead time, is four to eight days, including non operating time for the factory. The factory's operating time is from 6 a.m. to 10 p.m. Monday to Thursday and 6 a.m. to 4 p.m. on Friday's.

The scheduling also takes into account special characteristics of the window. If the window has two distinct colors or aluminium cladding it is marked in the bar chart with distinct colors, as presented in Figure 10. This is because these products usually have lengthier lead times. Gilje makes an effort to avoid producing too many unusual products at once because they occupy more capacity. The same applies for urgent orders, but in this case they have deposited additional capacity each day for these kind of orders, to prevent disruptions in the original schedule.

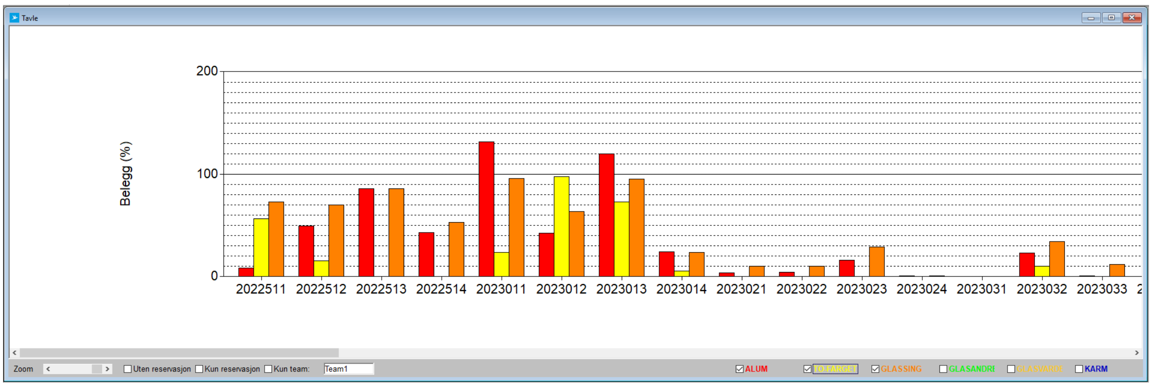


Figure 10: The bar chart from Gilje's production with filtering options for special characteristics. Red - aluminium cladding, yellow - multi-color, brown - regular order

The production priority of orders are derived from the bar chart as well. When the orders are placed in the bar chart they receive an unique order id, which determines the production sequence which ranges from lowest to highest order id. The orders planned to be produced the same day receive a coherent series id as well. The company follows a first-in first-out (FIFO) priority rule. The required glass panes for the orders are purchased from a supplier four days in advance. The production planning process from order acceptance to production starts is presented in Figure 11.

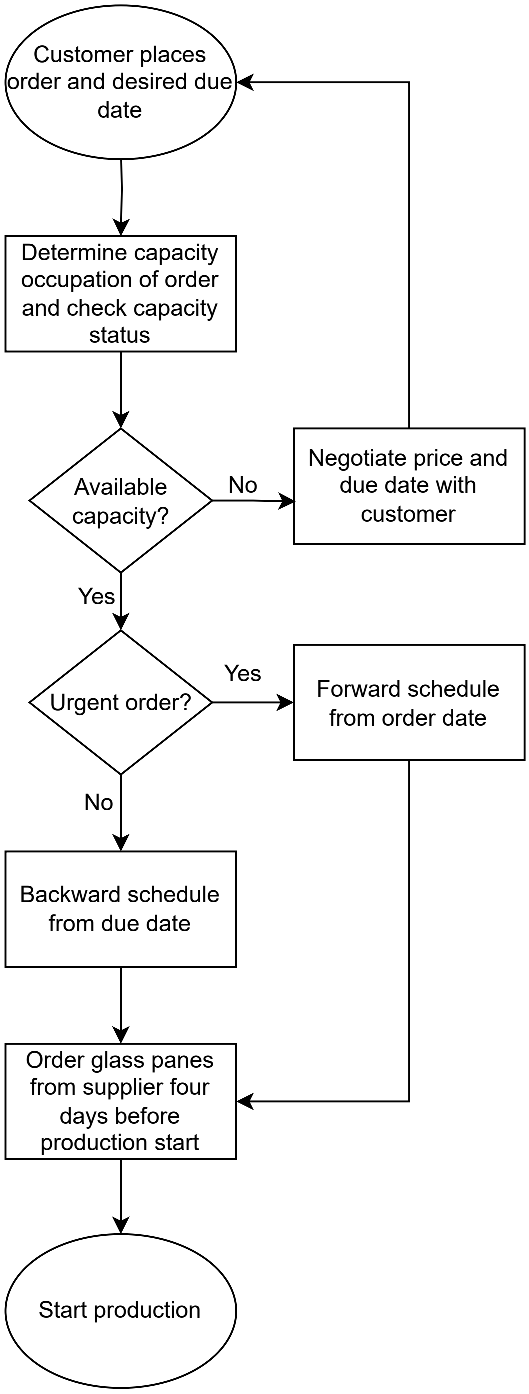


Figure 11: Flow chart of Gilje's order acceptance and production planning procedure

4.3 Production process

The manufacturing shop floor has a flow shop line layout. Gilje focuses on standardizing the production process, and almost all components have the same routing, independent of their processing needs. The only deviation is a separation between casement and frame components in some of the work areas. Defect components or components with special needs, like *Skråtapping* (bevel cutting), may also be taken out of the production line and processed manually in a remote working area.

Gilje's production process is presented in Figure 12. Firstly, the components are divided in separate lines for casement and frame where they are profiled, drilled and milled. After the individual shaping of the components they are gathered on the same belt. The components follow a FIFO sequencing rule through impregnation and priming. The process is fully automated until the buffer area. In the buffer area, before top coating, the components are manually placed in pallets with the corresponding components for a given window, and sorted according to color code. The sequencing for top coating is colored based. Light colors are processed early in the day, and darker colors later. Thereafter the casement and frame are separated through packing before the final assembly line. The frame, casement and glass panes are assembled, and optional aluminium cladding is applied before the final packing and shipping.

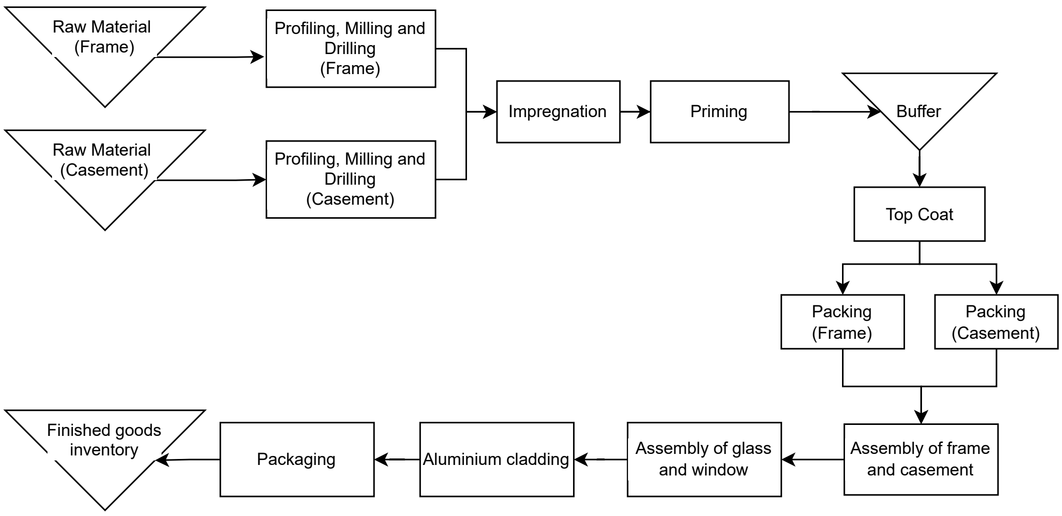


Figure 12: Gilje's production line

It is hard to obtain a standardized production process due to all the product varieties that occur. Gilje's customers can choose between over 2000 colors, and it is optional to have different colors inside and outside, or no color at all. The height and width of the windows vary from 25 cm to 7000 cm, and the customer can choose between 400 types of glass. The shape of the windows are also customer specific.

The production process suffers from two main waiting areas. The first is located between the impregnation and priming station where it may appear queuing time. Between these two stations there is located a conveyor belt which transports the components, see Figure 13. The components are allocated next to each other on the conveyor. The length of each component will therefore determine how many components the conveyor can transport in parallel. Since there is not utilized a sequencing rule taking the length into account, the number of components transported on this belt varies. Which causes queuing time for the subsequent components. Additionally, each component undergo the priming machine two times, causing longer processing time at the priming station than the impregnation station. This results in stops at the conveyor belt and queuing time for the components.

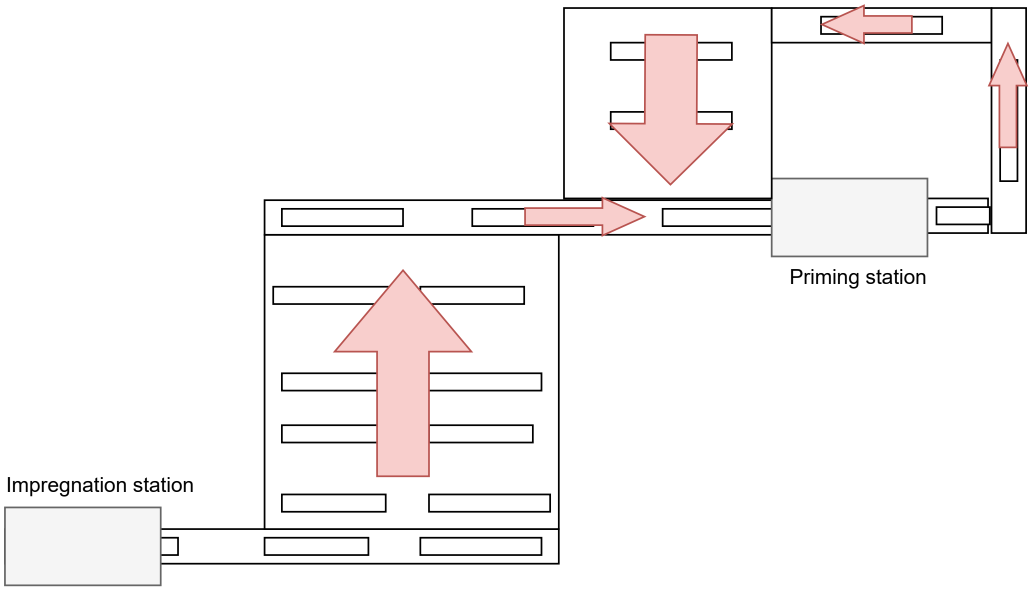


Figure 13: The material flow between impregnation and priming station

The second waiting area is located at the buffer before top coating, which Gilje stated as the bottleneck and the largest buffer/WIP in their production. Due to the sorting rules applied in the buffer the components may wait a long time. During last year, Gilje introduced a sorting approach according to colors in order to reduce the changeover time at the coating machine. Which they have expressed that they are satisfied with. The other aspect that causes waiting in this area are defect components, which have to be are reordered or manually processed. This sinks the lead time for the whole window, since the other components in the window needs to wait at the buffer zone until all components are ready for top coating. The third waiting area is located at the last assembly line, due to operators varying work pace.

4.4 Today's utilization of RFID system

Gilje has implemented an RFID system in their production line. The RFID scanning points at Gilje's shop floor is visualized in Figure 14. The RFID system at Gilje tracks the components at the shop floor, and the data can further be utilized in real-time or stored for later usage. An example of an RFID tag and a scanner is visualized in Figure 15, which are photos taken at Gilje's factory. As seen in Figure 14, three of the machines have scanners placed both before and after the machine. The RFID system is an important data source for this thesis.

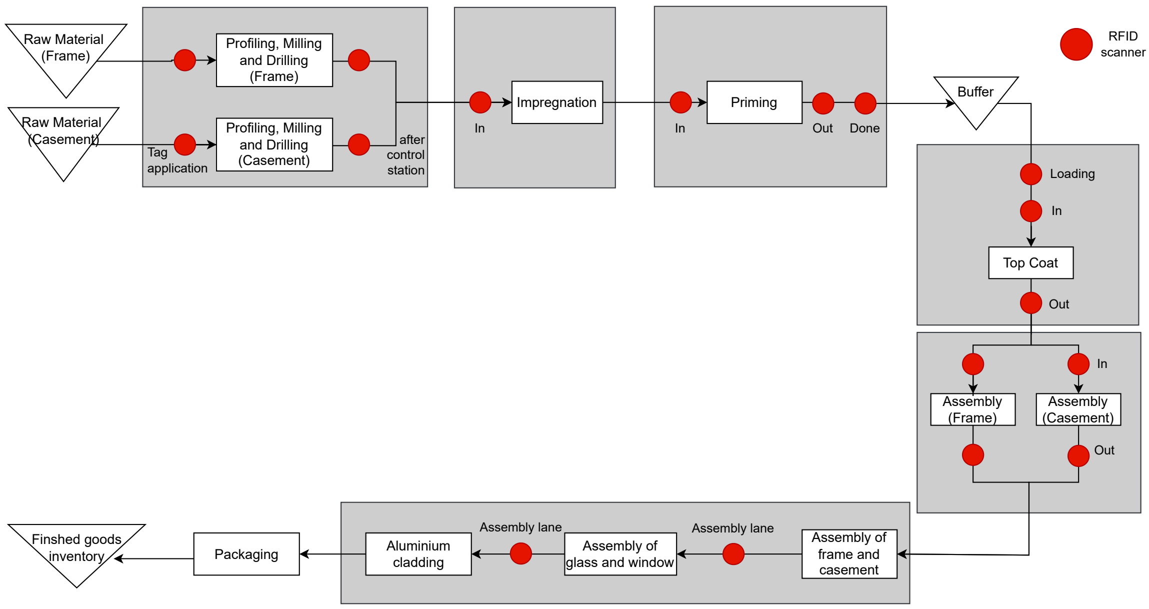


Figure 14: RFID scanner points in Gilje's factory

Through meetings, a visit to Gilje's factory at Gilja and interviews, the authors have identified how the company uses the RFID system today. They use the data to update the operators on which components that have been reordered through screens on the shop floor. The screens provide information such as the station from where the component was reordered and the time since the reorder, as shown in Figure 16a. They may also monitor the amount of daily production and the number of orders still remaining to be produced that day, as shown in Figure 16b.

The only place in the production line where the RFID data is utilized to automate the production is at the packing machines. The RFID data is read from the tag and inform the machine how the packing should be performed. Additionally, a test implementation for the priming machine has been conducted, indicating to the machine how many cycles of priming a component requires. Because the shop floor operators could not see the benefit and were content with the manual process, this was not put into practice.

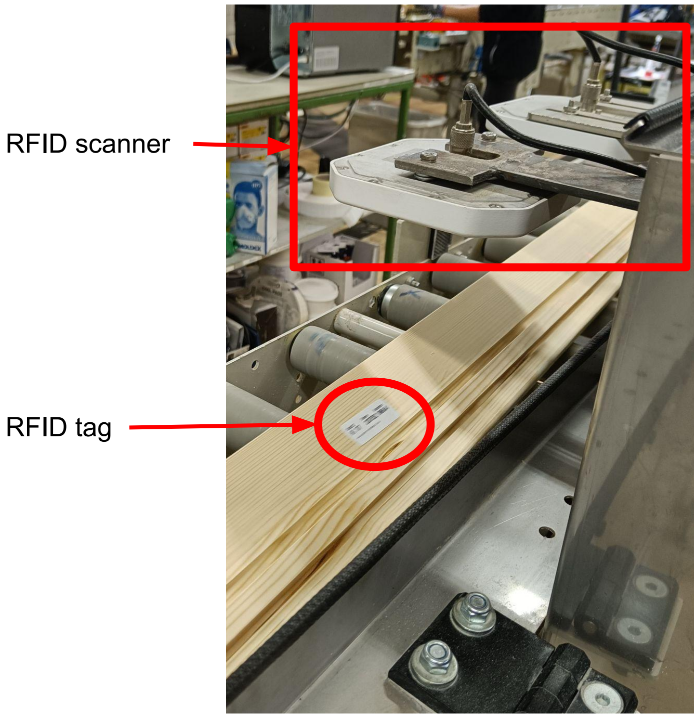
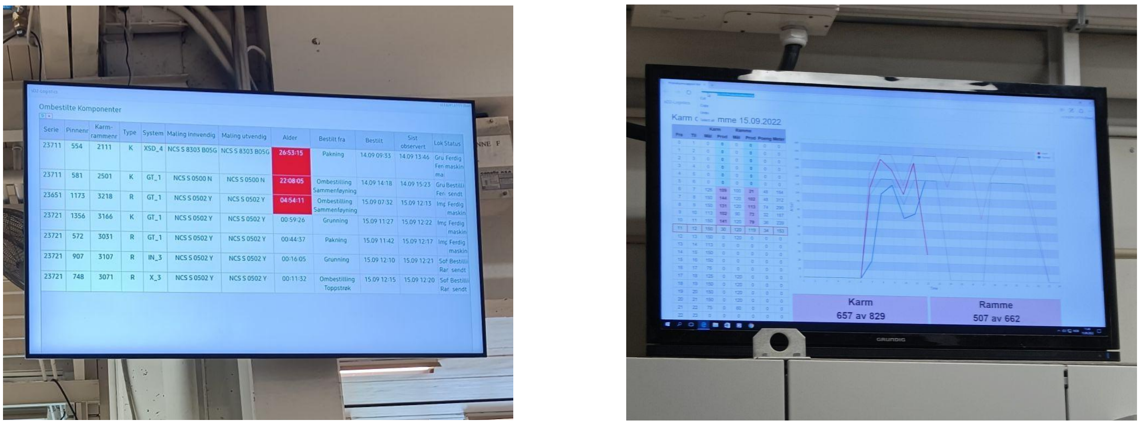


Figure 15: Photo of an RFID scanner and a component with RFID tag at Gilje's shop floor.



(a) Screen for reordered components.

(b) Screen for daily production level.

Figure 16: Screens at Gilje's shop floor that use RFID data

4.5 Analysis of today's planning approach

An analysis was carried out from Gilje's order data and bar chart data in order to obtain an overview of Gilje's delivery precision and glass point accuracy. The only measurement of their planning method today is derived from intuition and observations. Gilje stateed that there is ±10% deviation from the original production schedule in the bar chart. The company added that they have a delivery precision of 99 % as of today.

In order to assess Gilje's delivery precision, the analysis first investigated discrepancies between expected and actual delivery dates obtained from the order data. Table 7 displays the amount and percentage of orders that were delivered on time, ahead of schedule, or behind schedule from the order data. In Figure 17 the deviation from the original delivery date are shown for four randomly chosen dates.

Planned delivery date vs actual delivery date	Early delivery	Tardy delivery	Precise delivery
amount	3982	5282	8096
percentage	22.94%	30.44%	46.64%

Table 7: Deviation from scheduled delivery date

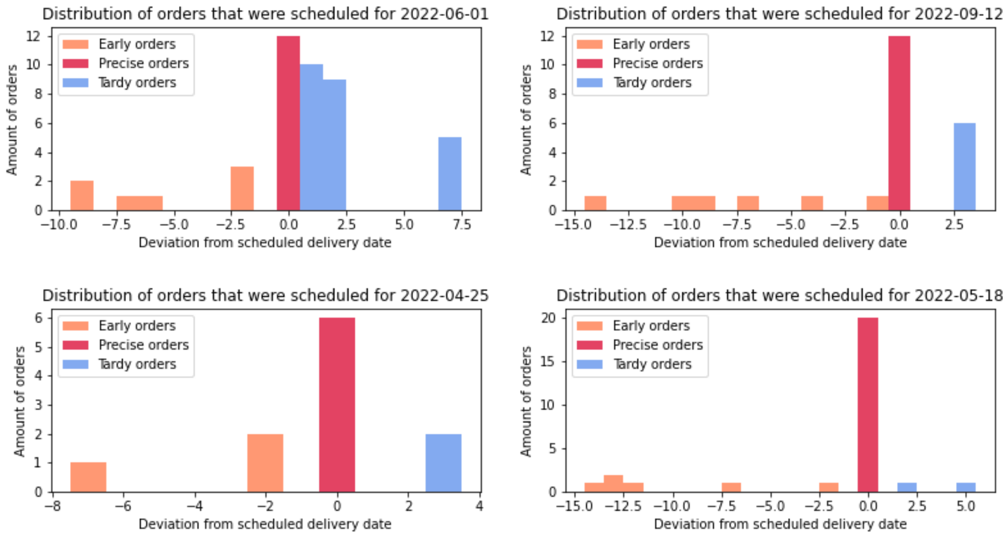


Figure 17: Deviation from scheduled delivery date distribution for four days

The results from both the table and the figure shows that there are several deviations from the original schedule, which contrasts with the statement of a 99% delivery precision. Only 46.64% of the orders where finished on the planned date. 3982 orders, which constitute 22.94% of the orders, where finished earlier than scheduled, while 5282 orders, constituting 30.44% of the orders, where finished later than expected. A precise estimation of the delivery precision is hard to obtain from the data since, as Gilje stated, they can change the delivery date in communication with the customer without updating the data records, therefore the real delivery precision might be higher than what the data shows.

The delivery precision on a weekly and monthly basis are presented in Figure 18. Data set 3 is limited to week 18 to week 41, the analysis will therefor only take into account this time frame. Week 29 to week 31 are also excluded since Gilje are not operating in this time period.

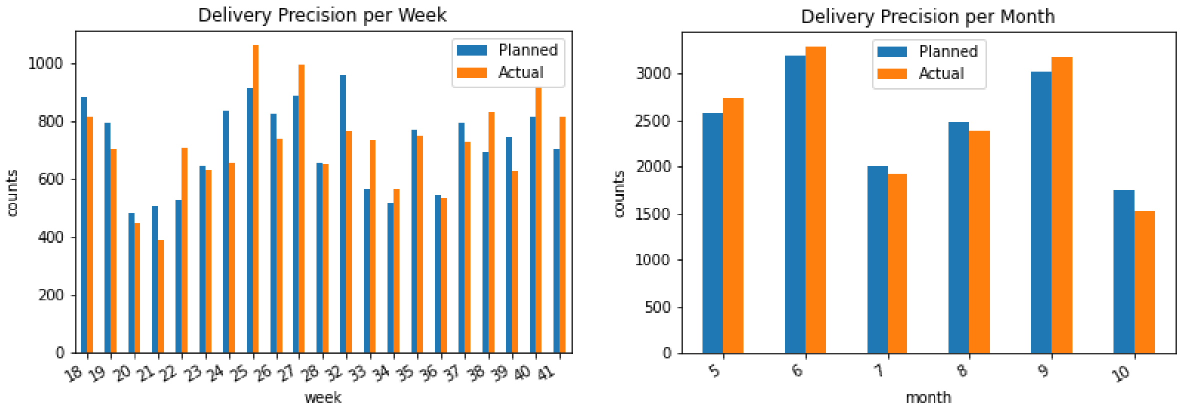


Figure 18: Delivery precision on a weekly and monthly basis

Secondly, the glass point precision was investigated from the bar chart data. The datasets provided information regarding anticipated and actual amount of glass points, which was compared. The results are displayed in Figure 19. The figure shows both the glass point precision on a weekly- and monthly basis. Both graphs shows deviations between the planned amount and actual amount.

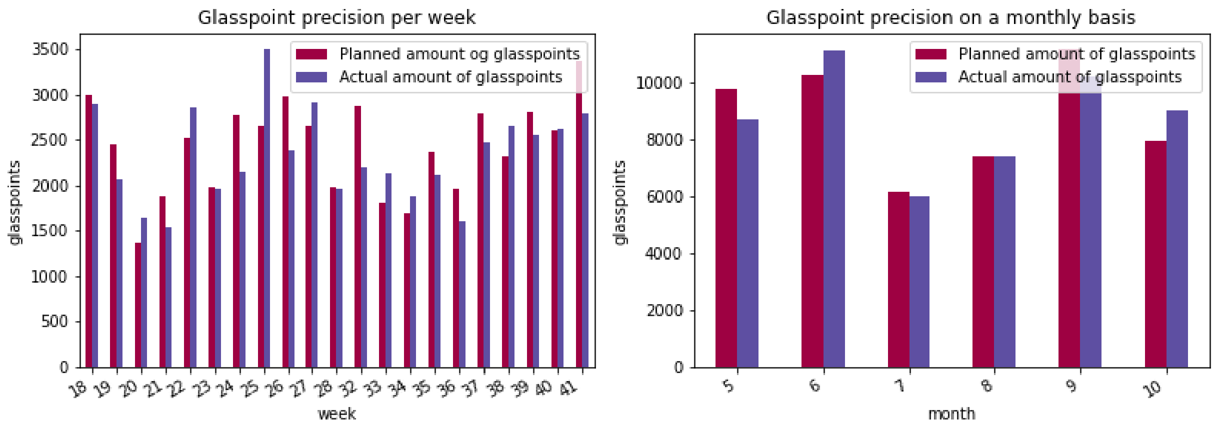


Figure 19: Glass point precision on a weekly and monthly basis

4.5.1 Opportunities for an improved planning approach through lead time prediction

The analysis showed that Gilje has improvement potential in order to obtain a more precise planning system. Gilje has stated that reduced delivery precision may be due to exceeded production capacity, which can lead to production delays and deviations from the original plan. Gilje's capacity planning is today based on one order characteristic, number of glass panes. It does not take the shop floor load or other dynamic variables into account, despite the fact that they have implemented an RFID system which provides the necessary data to do so.

The authors and company therefore see a potential of adapting LTP in Gilje's planning system. Today the lead time is solely estimated based on intuition and observation, and the same lead time is assumed for all orders. If the order is completed sooner than expected, it is stored early, causing a higher finished goods inventory, further increasing inventory costs and deviations from the original schedule. If the order is completed later than expected, it can cause congestion in the production line, and the company can fail to meet the due date, causing poor customer satisfaction. In order to prevent the latter, Gilje has added the buffer of five days after the completion day. These extra days are useful in case of abnormalities, but they can also cause unnecessary inventory goods, inventory holding cost and waiting time for the customer.

The RFID data makes it possible for Gilje to track and receive relevant information about each component, as well as assessing the status of the production system in real time. The available data provides useful information about the system status, that can further be used to estimate a precise lead time which can be incorporated in Gilje's planning method. In addition to the RFID data, Gilje has access to order data from their ERP system and glass point data for the planned orders in their bar chart.

In order to achieve business understanding for the task at hand, the necessary input features for LTP were discussed with the company. As mentioned in section 2, the case study is a continuation of a previous work conducted by the authors. The lead time affecting factors detected from the literature were discussed with Gilje, as well Gilje's thoughts about additional factors were identified in the prior study through a questionnaire and meetings. Gilje mentioned bottleneck and WIP/buffer as the most crucial factors since it may cause production delays and halting. In their case the buffer before the top coat is recognized as the potential bottleneck. Additional waiting time is also given a high weight of importance. Product mix, utilization and machine downtime of the machines, changeover time, type of product, number of components, number of glass panes, color, and number of colors are additional factors that Gilje underlined to be important for the lead time.

5 Data analysis

This section presents the applied data analysis approaches. Firstly, the utilized methods for data collection and preparation is explained and justified, in section 5.1. Thereafter, the modelling and development of the chosen ML models are presented in section 5.2.

The main Python pipeline for the project is presented in Figure 20. Firstly, the data provided by Gilje was retrieved. The data was further explored and assessed through a data quality assessment. Detected faults were removed during data preparation, and system status features were calculated. Thereafter the dataset was split into training and test set resulting in a complete input dataset for the ML model. This study developed two input datasets, with and without system status. The selected ML models, RF and MLP, were compared to a baseline performance, a linear regression model. Additionally, the two input datasets were compared to see the effect of including system status variables. Finally, hyperparameter tuning was applied on RF and MLP, which gave two final models for comparison.

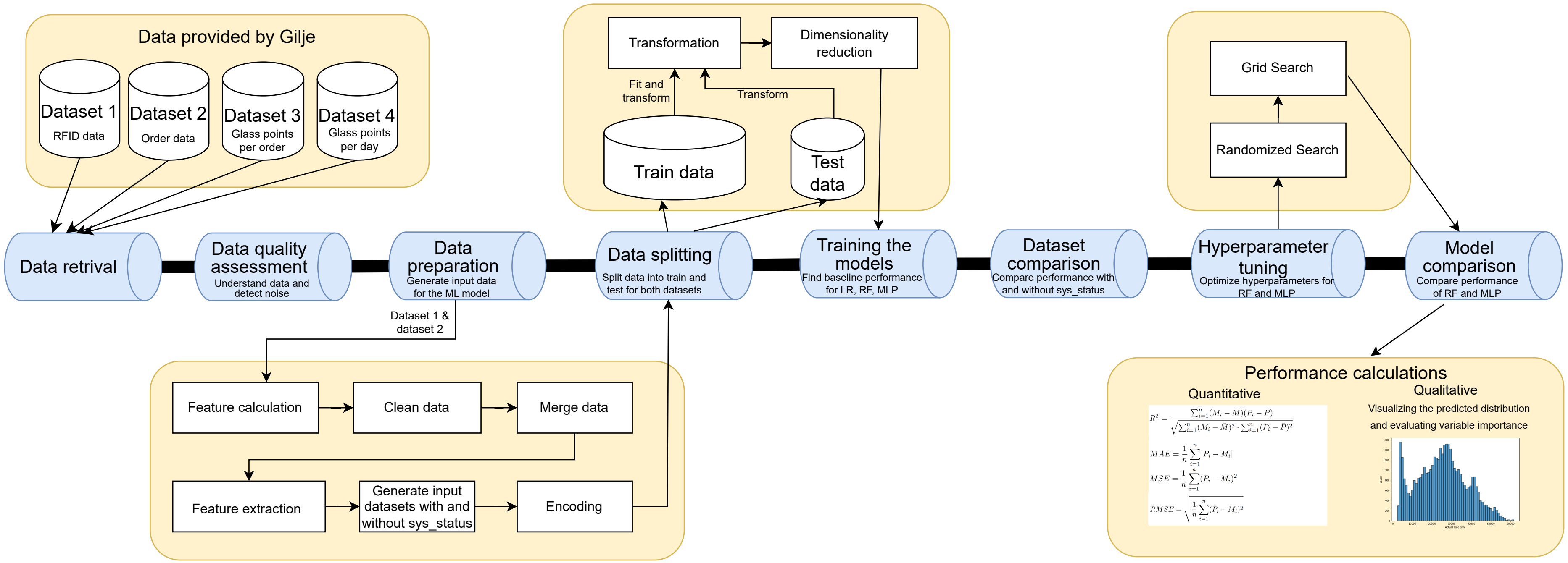


Figure 20: Main python pipeline of the work

5.1 Data acquisition

This section contributes to the second and third step of CRISP-DM, data understanding and preparation. In order to obtain a data understanding the data was collected and examined. In section 5.1.1, the collected data is presented and explained. Further, in section 5.1.2, the applied preparation techniques are explained, resulting in two input datasets for the ML models.

5.1.1 Data collection

The case company, Gilje, has provided data to the study. Four datasets were utilized which Gilje derived from their database. The data has been reviewed in corporation with Gilje through interviews and discussions. The datasets provide information about shop floor events, order data and planning data. The four datasets are summarized in Table 8 at the end of this section. The data was collected over a time horizon of about six months. All data was received in structured or semi-structured tables in xlsx or json format.

The shop floor data was obtained through Gilje's RFID system, referred to as dataset 1. Gilje's RFID system store information about component characteristic and timestamp tracking throughout the production. All components that enters the shop floor has an RFID tag attached, and every machine has an RFID scanner before and/or after the component enters or exits the machine, as explained in section 4.4. All components in the dataset had time records from each RFID scanning point the component passed. This study did not receive access to real-time data due to accessing restrictions, and it therefore relies on historical data. Dataset 1 contains records from 19.04.2022 to 17.10.2022. The data is not continuous in the whole time period due to downtime in the production during summer vacation in week 29 to 31, and some days during the Easter holiday.

Data about the manufactured orders was retrieved from Gilje's ERP system, referred to as dataset 2. This data gave further insight to the order characteristics. The data records included information about each window in an order. Information about the customers and characteristics of the windows, e.g. color, length, type of frame and casement etc., was included in this data. Additionally, the data included actual due date and planned due date for each order.

The remaining two datasets were derived from Gilje's bar chart, and constitute relevant planning data. They provided details about the manufacture's projected production. The first dataset describing planning data, referred to as dataset 3, contained the number of glass points allotted to each order. In contrast to dataset 2, dataset 3 contained records for each order, which often includes several windows. The fourth dataset, known as dataset 4, contained records about the entire amount of glass points each day that were available and scheduled. This comprise the total number of glass points assigned to all orders planned to be produced a specific day. The data did not specify which orders that were included each day.

Name	Description	Original dimension
Dataset 1	RFID data from the production line	287 014 rows, 50 columns
Dataset 2	Order data from Gilje's ERP system	24 787 rows, 23 columns
Dataset 3	Amount of glass points assigned to each order	7855 rows, 5 columns
Dataset 4	Amount of available and planned glass points each day for the production	115 rows, 12 columns

Table 8: The original datasets Gilje provided to this study

The dimensions of the datasets varied greatly, as can be seen in Table 8, despite being retrieved in roughly the same time period. As previously stated, a customer order, which is recognized by an order id, may contain several windows. A window is made up of several components. Dataset 1 contained records about components, and therefore had several rows compared to the other datasets. The records in Dataset 2 were based on the windows in an order, and each window was listed in separate rows. Dataset 3 contained records of each order, based on their order id. Dataset 4 had one row per day and showed tracking of the bar chart over 115 days. The time periods in the datasets did not entirely overlap. The tracking duration for dataset 1 was 181 days, and the tracking period for dataset 4 was 115 days. Non-overlapping time periods were therefore excluded.

The four datasets were all evaluated as potential inputs for the ML model. However, dataset 1 and dataset 2 were found most appropriate for this task, since they contained information about system status and order characteristics. From these datasets, the authors were able to extract valuable input features that matched the input features detected through the case study and literature review. The amount of glass points for each window was derivable from dataset 2, and inputs from dataset 3 were thus not required. Dataset 4 was not usable as input for the ML model since the data lacked records of the order id.

5.1.2 Data preparation

In this section, the applied data preparation techniques on dataset 1 and 2 to achieve a final input dataset for the ML model is presented. This included the steps of feature calculations, data cleaning, feature extraction, encoding, train-test split, transformation and dimensionality reduction, summarized in Figure 21. The data cleaning was based on the detected problems during the data quality assessment, presented in section 6.2.

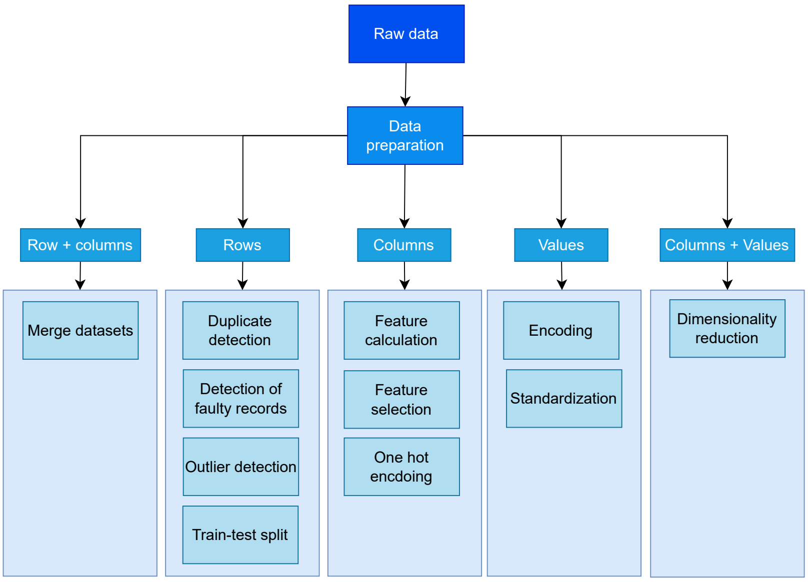


Figure 21: Overview of applied data preparation methods.

Feature calculation

The literature has identified system status as an important factor affecting the lead time. None of the datasets contained attributes of system status variables. However, several variables were possible to calculate from the RFID time stamps in dataset 1. Therefore the shop floor load, buffer load, corresponding to the systems WIP, and amount of components in a series were calculated at the time the component entered the production. Additionally, the lead time for each component was not provided as a feature. Since supervised ML models are target based, and the desired output needs to be included in the input dataset the components lead times were also calculated based on the RFID data.

The lead time for each component i (LT_i) was calculated by deriving the first event time for component i (ET_i^{first}) and the last event time for component i (ET_i^{last}), where ET_i^{first} and ET_i^{last} are the first and last RFID scanning records of the components at the shop floor. WD_i^{non} is the time component i spends in the production when the manufacture is not operating. The formula presented below was applied for each data record in dataset 1:

$$LT_i = (ET_i^{last} - ET_i^{first}) - WD_i^{non}$$

The shop floor load of component i (SFL_i) was found by evaluating the first event time of component i (ET_i^{first}), which was checked to be after the first event time of component j (ET_j^{first}) and before the last event time of component j (ET_j^{last}), for all j in the dataset (d):

$$SFL_i = \sum_{j=0}^{j=len(d)} (ET_j^{first} < ET_i^{first}) \,\&\, (ET_j^{last} > ET_i^{first})$$

The buffer load between priming and top coat for component i (BL_i) was derived in the same way. Where $ET_j^{priming}$ is the last event time at the priming station for component j, and $ET_j^{top\_coat}$ is the first event time at the top coat station for component j:

$$BL_i = \sum_{j=0}^{j=len(d)} (ET_j^{priming} < ET_i^{first}) \& (ET_j^{top\_coat} > ET_i^{first})$$

The last parameter that was calculated and added to the final dataset was the amount of components belonging to the same series as a component i, namely the *series_amount*. The series amount was calculated for all unique series, and the corresponding *series_amount* of component i was added as a feature based on it's series number.

Data cleaning

Firstly, Dataset 1 was independently assessed in order to detect and remove inaccurate data records, which would have been noise for the ML model. All components that had been manually processed or reordered were eliminated. These components were removed in the dataset due to inconsistent scanning points, which resulted in unusually long lead times since the components had exited and reentered the production line in a non-consistent order. Data attributes such as quality check status and rework status were regarded as irrelevant as they are not retrievable at the time the order enters the production.

Duplicates and missing values are noted as crucial issues to remove in order to achieve a reliable dataset (Maharana, Mondal et al. 2022). 14 449 duplicates were found in dataset 1, which were removed. Some features were removed due to homogeneous values for all data points in the dataset. Consequently, they don't offer any information that the ML model could use. This concerned the features *type of wood* and *skråtapping*. However, if the data was captured over another time period, these features should be re-checked since they then might contain other values.

Several flawed data records were detected through an analysis of the RFID time stamps. Figure 14, presented in the description of the case study, display the RFID reading points that a component passes during typical operational conditions. This contains eleven reading points for the frame and eleven reading points for the casement prior to the assembly line. In the raw data some components had over 20 000 reading points, and several components only included some of the reading points while leaving out others. To ensure that the components included in the input dataset had consistent tracking on the shop floor, all components that did not pass all eleven shop floor reading points were excluded.

The distribution of lead times for all components in Dataset 1 is presented in Figure 22. Several outliers were found from analysing the lead time distribution, specially in the upper quartile. The outliers were removed since it is likely that they have appeared due to scanning errors.

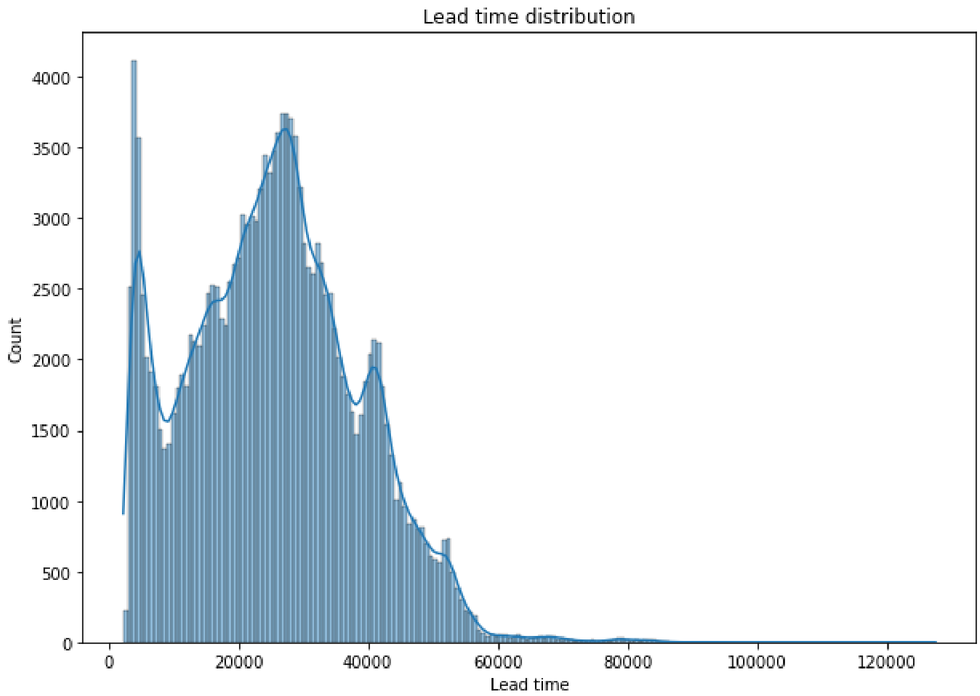


Figure 22: Histogram of components lead times

The interquartile range (IQR) method was applied on the lead time feature to detect and remove these outliers. The method removes values in a left or right skewed distribution, and from Figure 22 it can be seen that the lead time distribution was right skewed. The IQR method divides the data into 25% quartiles, Q1, Q2 and Q3. The formula is given below, and iterates through a recursive algorithm starting from the median (Vinutha, Poornima et al. 2018):

$$IQR = Q3 - Q1$$

$$Lower\ boundary = Q1 - (1.5 * IQR)$$

$$Upper\ boundary = Q3 - (1.5 * IQR)$$

The distribution of the lead time features before and after applying IQR is presented in Figure 23. As can be seen in the figure, the right skewed distribution was eliminated after utilizing IQR.

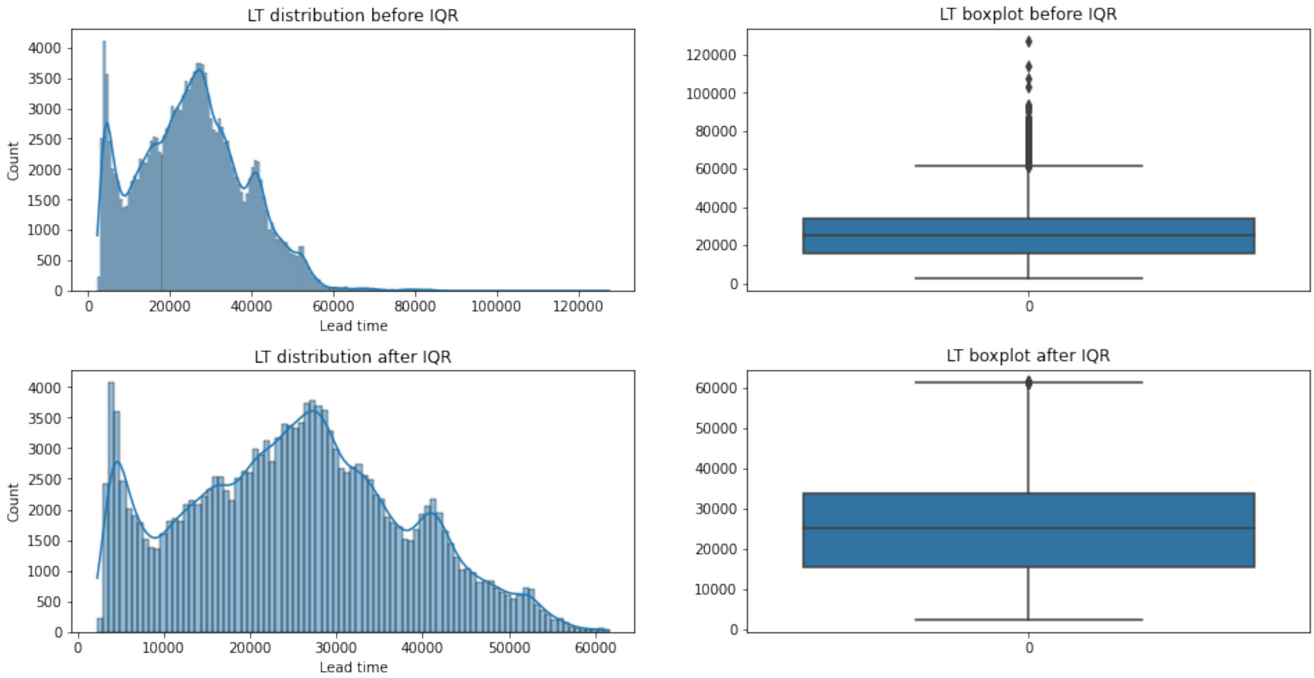


Figure 23: Lead time distribution before and after applying IQR outlier detection

Dataset 2 was also individually assessed before the datasets were joined. In communication with Gilje, all records in the 'lev eno uno' column without the value 'UNO' were deleted. They were eliminated because the orders require special handling, which does not occur under normal production. Fastlane orders were also removed since these orders are scheduled uniquely with an independent capacity buffer.

The two datasets were then merged to a coherent set of data, referred to as dataset 5. Dataset 1 and 2 were merged on order id and row number. The order id represent which order the component belongs to. An order usually consist of several windows and the windows are separated with a unique row number. The merge secured that dataset 5 only consisted of orders and components that appeared in both datasets. Additionally, each component have an unique component id which was used to separate the components belonging to a window. The remaining data included 128 888 rows. The final statistical description of the lead time is shown in Table 9.

Statics	**Value**	**Description**
Count	128 888	Number of components
Mean	25 053 sec	Average lead time in seconds.
Std	12 577 sec	Standard deviation of the lead times.
Min	2 278 sec	The shortest lead time occurrence.
Max	61 610 sec	The longest lead time occurrence.
25%	15 652 sec	25% of the data has a lead time of 15 652 sec or lower.
50%	25 149 sec	Median - 50% of the data has a lead time of 25 149 sec or lower.
75%	33 549 sec	75% of the data has a lead time of 33 549 sec or lower.

Table 9: Statistical characteristics of the lead time feature after data cleaning

Feature extraction

After the merge of dataset 1 and 2 the number of features needed to be reduced by removing overlapping and irrelevant information. The original datasets contained a large number of features, with 53 features in dataset 1 and 23 features in dataset 2.

Given the size of dataset 5's feature space, the authors saw it suitable to manually assess each feature's significance before applying methods for feature selection. The foundation for the feature extraction was achieved through the questionnaire and meetings with Gilje, presented in section 4.5.1, and through findings from the literature. Several attributes were removed since they did not affect the lead times, and therefor added noise to the ML models rather than useful information. The remaining features in dataset 5, which were utilized as final inputs for the ML model, are presented in Table 10.

Name	**Description**
prods_belegg	The amount of glass points given to the specified window
ant_karmkomp	The number of frame components belonging to the window
ant_rammekomp	The number of casement components belonging to the window
maling_utv	The color code for the external painting
maling_innv	The color code for the internal painting
maling_utv_inv_diff	Specifies if the external and internal painting are different colors
ant_fag	Specifies if the casement is made horizontally or vertically, and if there is need for a frame or just the glass
length	The length of the component
width	The width of the component
thickness	The thickness of the component
grunning	Rounds of priming for the component
runder_toppstrøk	Rounds of topcoat for the component
dyser	If the component have nozzles on or off
type_karm	The casement type for the component
type_comp	The frame type for the component
brann_klassifisering	If the painting has a fire classification
sparkleinfo	Information about necessary plaster
nedtapping	If the component has draining needs
series_count	Amount of components belonging to the same series
count_comp_level	The number of components at the shop floor when the order enters the system
count_buffer_level	The number of components in the buffer between priming and top coat when the order enters the system

Table 10: Input features for the ML models

Data encoding

A ML model is not capable of understanding free text, and it was therefore necessary to encode categorical features. In simple words, the data points were converted to values of 0 or 1, dependent on the investigated value of the feature category (Maharana, Mondal et al. 2022). One-hot encoding was the applied encoding technique, which encodes a feature by creating a new feature for each distinct data point that appears in the initial feature collection, as shown in Figure 24. This technique was applied due to it's simplicity and interpretability.

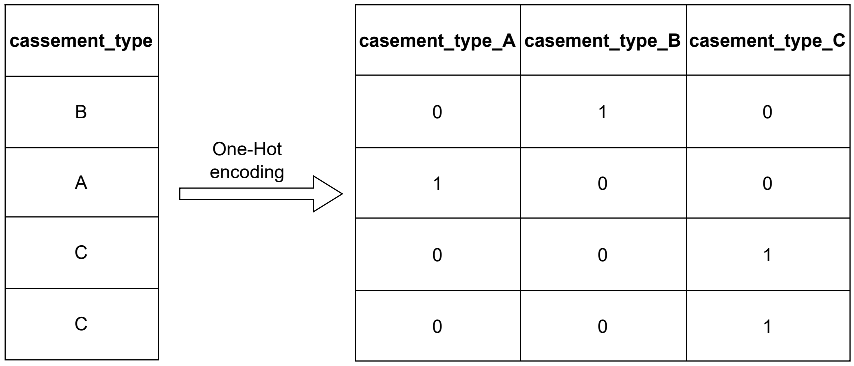


Figure 24: One-Hot encoding of categorical features in the dataset.

The majority of the categorical features were subjected to one-hot encoding, while some of the columns required additional preparation. The feature *brann_klassifisering* only needed to be categorized according to whether it had a fire classification or not. If one-hot encoding had been directly applied, all unique fire classes would have received a separate column which was not desired. Therefore, each value in the *brann_klassifisering* column were converted to either 1 or 0 depending on whether the original value was a fire classification or not. The same applied for the painting-related features. Since there were more than 150 distinct color codes in the dataset. The colors were encoded by separating them between dark and light colors, as Gilje does in their production. Thereafter, one-hot encoding was applied.

Train-Test split

The main objective of any supervised ML model is to perform well on new data, which the model has not seen before. This necessitates training and testing the machine learning model on separate datasets. Therefore the data is separated into train and test sets.

The *training set* is applied in order for the model to learn the relationship between the input features and the target variable. This is the dataset that the model is trained on, and it is allowed to see both the input and output pairs of variables. The training set consisted of 67% of the original dataset.

The *test set* is used to evaluate how well the model performs on new unseen data. The model has only access to the input features, and needs to predict the output variable based on the relationship learned from the training set. The difference between the actual target variable and the predicted variable is used to evaluate the performance of the model. In this thesis 33% of the data was used as test data.

Data transformation

Data transformation was necessary since the features' values appeared in different scales. For instance, the length of a component was in meters but the lead time was in seconds. The numeric value of lead times were likely to be greater then the numeric values of the length, and the model would have given more weight to the lead time variables if the data had not been transformed. Therefore standardization was applied in order to achieve a weight according to natural differences and not experimental artifacts (Muralidharan 2010).

Standardization calculates the mean, μ, and standard deviation, σ, of all the values of a feature, z. For each data point, x_i, the mean is subtracted and divided by the standard deviation. The outcome is a mean of 0 for all values and a standard deviation of 1. The formula is given below:

$$z = \frac{x_i - \mu}{\sigma}$$

Standardization was performed after the dataset was divided into a training and a test set. The training data was fitted and transformed in order to learn the parameters of the scaling and to scale the data. The test data was only transformed since the parameters of the data learns the scaling from the training data.

Dimensionality reduction

Dimensionality reduction tools were applied, after manual feature extraction, to reduce the feature space even more. Pearson correlation coefficient was used to detect the correlation between the features, and principle component analysis was used to reduce the dimension. Maharana, Mondal et al. (2022) notes that reducing the dimension of the data can increase the models performance.

Pearson correlation coefficient

The Pearson correlation coefficient (PCC) assesses the degree of correlation between two variables, as well as the direction of the correlation (Benesty, J. Chen et al. 2008). The values of the two variables travel in opposition to one another if the correlation value is negative. The values of the two variables move in the same way if the correlation value is positive. Highly correlated variables can sometimes contain redundant information, causing noise in the dataset. The PCC can be used to identify these inter-related variables (Jayaweera and Aziz 2018). The mathematical formula is given below, where X and Y represent two different variables (Dalinina 2017):

$$p_{X,Y} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

The PCC analysis was performed on the numerical features in the dataset in order to visualize possible correlations between them. The correlation matrix is displayed in Figure 25. Most of the attributes had low correlations according to the Pearson coefficient. However, some coefficients indicated a strong relationship, and some features had a 100% positive correlation. The findings showed that there existed interrelated variables with the potential to hold redundant data, and dimensionality reduction was thus of interest. For example, the three features *grunning*, *dyser* and *runder_toppstrøk* were all perfectly positive correlated, meaning that the features moved simultaneously by the same amount and in the same direction. These three features contained redundant information, and could further be reduced to one representative column. Additionally, there was a relatively high correlation between *prods_belegg*, *ant_karmkomp* and *ant_rammekomp*, where the correlation moved in a positive direction. This can also be seen for the three system status features, *count_comp_level*, *count_buffer_level* and *series_count*. It appeared from the PCC analysis that a number of features were connected, and as a result, a principal component analysis was applied to eliminate redundant information.

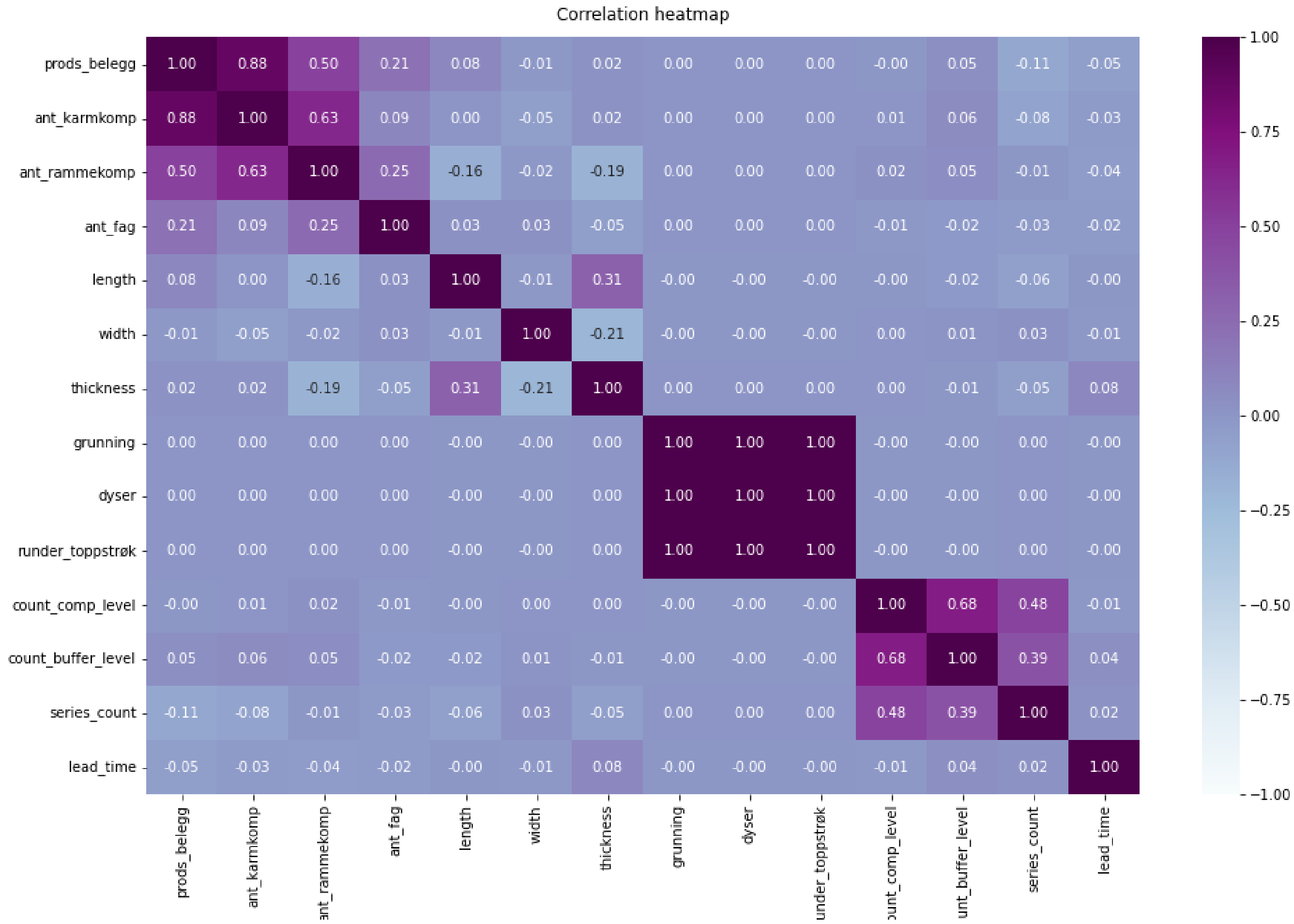


Figure 25: Pearson correlation coefficient plot

Principle component analysis

Principle component analysis (PCA) is a technique used to reduce the dimensionality of the data (Maćkiewicz and Ratajczak 1993). Through PCA a new set of uncorrelated features are derived based on the original features (Hasan and Tahir 2010). These new features are called principle components (PC), and the aim is to find the fewest number of features that are able to capture the most relevant information from the data, retaining only those features that have a significant amount of variance.

There are several approaches when determining the final dimensionality of the dataset based on PCA. This paper considered three different approaches. Firstly, *Cumulative variance* was applied, for determining the amount of PCs to retain. This approach is based on the size of the eigenvalues or the percentage of the variance described by each PC. *Kaiser Gutman Rule* was the second approach, where only PCs with eigenvalues greater than one were detained. The final approach investigated was the *Scree plot*, which is a graphical representation of the eigenvalues on the y-axis and the PCs on the x-axis. The amount of PCs to keep are determined based on where the curve flattens. This is called the "elbow" point of the curve (Hasan and Tahir 2010).

The results from the PCA is shown in Figure 26. The plot to the left is the Scree plot. The elbow point was found after approximately 27 PCs. The plot in the middle is an extension of the Scree plot, where the Kaiser Gutman Rule is included. The horizontal red dotted line along the y-axis separates the components with eigenvalues smaller or larger than one. There were 17 PCs with eigenvalues larger than one. This suggested that 17 PCs should be kept in the final dataset. The graph to the right in Figure 26 shows the cumulative variance of the eigenvalues. In order to represent all variance in the dataset, the figure suggests to keep approximately 26 PCs based on the flattening point. Considering the different plots, the authors thought it was sufficient to keep 95 % of the variance, resulting in 23 PCs. The PCA was performed after the dataset was divided into a training and a test set, with the same procedure as for data transformation.

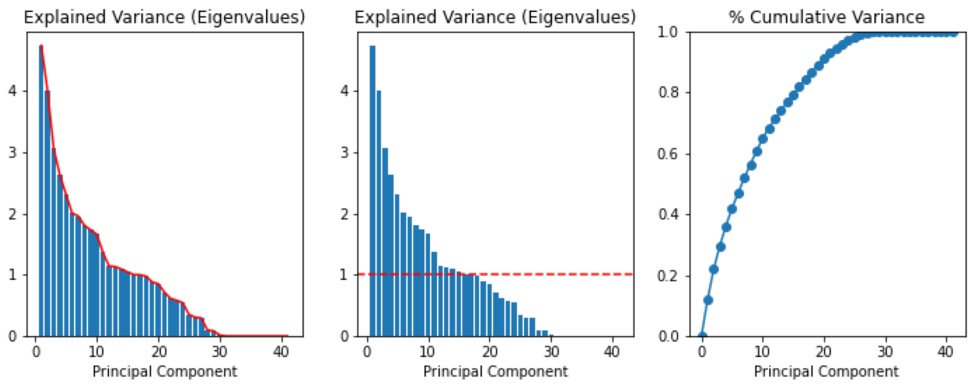


Figure 26: PCA results. From the left: Scree plot, Scree plot with Kaiser gutman rule, Cumulative variance of the eigenvalues

In summary, the applied data preparation techniques consisted of feature calculation, data cleaning, feature extraction, data encoding, train-test split, data transformation, and dimensionality reduction. The applied data preparation for the final input dataset is summarized in Figure 27. In order to investigate the importance of system status, the authors generated two final datasets. One dataset without shop floor load, series count and buffer load, referred to as *ml_dataset*, and another with all attributes, referred to as *ml_dataset_sys*.

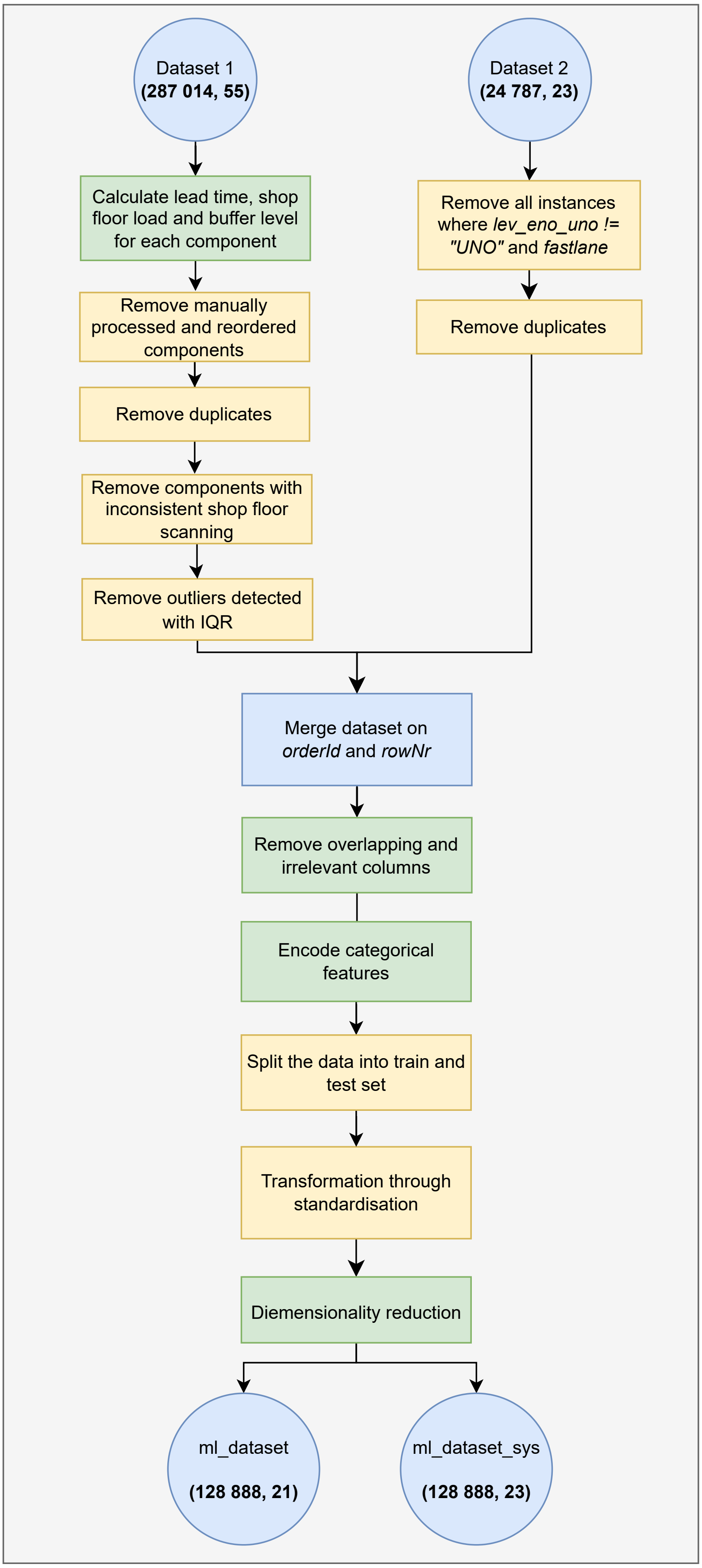


Figure 27: Overview of the required data preparation. Yellow boxes indicates changes in the rows, and the green boxes are changes in the columns.

5.2 Data Modelling

The two machine learning models utilized in this thesis are introduced in this section, along with evaluation criteria used to evaluate the results of the models. The section also provides an overview of the two hyperparameter tuning techniques that were used to find the optimal hyperparameter values in the study, known as grid search and randomized search.

5.2.1 Evaluation criteria

Every statistical evaluation measure only provides one interpretation of the model's errors, highlighting a particular feature of the error characteristics of the model performance (Chai and Draxler 2014). Therefore, in order to fully understand and evaluate the model, several performance indicators may be required. This thesis evaluated the performance of the ML models based on four types of performance evaluation indicators that are commonly used in the literature, namely coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE) and mean absolute error (MAE). The performance metrics are described below with the belonging mathematical expression. In the equations, n is the number of inputs, M indicates the actual value and P is the predicted value (Shah, Javed et al. 2021).

Coefficient of determination (R^2)

The performance metric indicates how much of the variance in the resulting variable that can be explained by the independent input features, and the value ranges between 0 and 1 (Shah, Javed et al. 2021). For example, if the R^2 provides an output of 0.6, it means that 60 % of the variance can be explained by the independent features. Higher R^2 values indicates higher dependency between the independent and dependent variables, meaning that the independent variables have a strong explanatory power.

$$R^2 = \frac{\sum_{i=1}^{n}(M_i - \bar{M})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n}(M_i - \bar{M})^2 \cdot \sum_{i=1}^{n}(P_i - \bar{P})^2}}$$

Mean Absolute Error (MAE)

MAE evaluates the difference between predicted and the actual observed value, and returns the average of the absolute values of all the residuals. Lower MAE values indicate lower errors and better model performance, and it is commonly used to measure the accuracy of regression models (Shah, Javed et al. 2021).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - M_i|$$

Mean Squared Error (MSE)

While MAE measures the absolute value of the difference between the actual and anticipated values, MSE takes the average of the squared discrepancies. MSE frequently has greater values than MAE since it penalizes larger errors due to the squared value (Chai and Draxler 2014). An improved model fit and fewer errors between the actual and anticipated value can be observed with a lower MSE value.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(P_i - M_i)^2$$

Root Mean Squared Error (RMSE)

RMSE measures the squared root of MSE. It is often more understandable than MSE, since it standardizes the measure units provided by MSE and return the value in the same scale as the dependent variable (Chicco, Warrens et al. 2021). In general, lower RMSE values are preferable because they show that the model's predictions are similar to the actual values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - M_i)^2}$$

5.2.2 Machine learning libraries

In order to process the data and generate the ML models, various Python libraries, including Scikit learn, Numpy, MatplotLib, Seaborn, and Pandas, have been applied in this study.

Scikit learn (Sklearn) is a python based ML library, featuring numerous state-of-the-art ML algorithms and techniques (Pedregosa, Varoquaux et al. 2011). In this study, several preprocessing methods from the library were used, including PCA, StandardScaler and pandas.get_dummies(), which is a version of Sklearn's OneHotEncoder. These methods were used to reduce the dataset's dimensionality, standardize the data, and convert the categorical values into numerical values. Various modules from the Sklearn library were also used in the model construction and model evaluation processes, see section 5.2.3 and section 5.2.4.

Fetching and managing the dataset was done using Pandas, which is a library used for data analysis (Bisong and Bisong 2019). The datasets were loaded using Panda's read_json() and read_excel() function, which converts the data from JSON or XLSX to a pandas Dataframe object. In the data exploration and data preparation phase, Pandas functionality was actively used to easily merge dataframes, add necessary columns and remove noise from the data.

Numpy is a python library used for optimization of numerical computations (Bisong and Bisong 2019). When working with higher dimensional data arrays, which is commonly required when utilizing ML, this library is widely employed. In the ML experiments, the test and training sets were represented using Numpy arrays.

For the visualization of the data, both Matplotlib and Seaborn has been applied. Through graphical visualization, data scientist are able to develop a higher understanding of the data and gain insight into the data structure (Bisong and Bisong 2019). Both libraries work well together with Pandas and Numpy objects.

5.2.3 Hyperparameter tuning

Hyperparameter tuning is an essential step in the modeling phase of the ML process, and it involves searching trough a grid of hyperparameter values, the hyperparameter space, and selecting the hyperparameter values that optimizes the model performance (Bardenet, Brendel et al. 2013). The Sklearn library offers two common techniques for searching the hyperparameter space, namely *GridSearchCV* and *RandomizedSearchCV*. Both techniques uses cross-validation to assess a model's performance and choose appropriate parameters (Pedregosa, Varoquaux et al. 2011). They select the best parameter values from a predefined parameter grid based on which hyperparameter values that provide the highest score (Pedregosa, Varoquaux et al. 2011).

Grid search involves exhaustively searching through all combinations of the specified hyperparameter values and retain the best combination (Bisong and Bisong 2019). However, this exhaustive search can be computationally expensive and time-consuming, especially when dealing with a large number of hyperparameters or a large search space (Valarmathi and Sheela 2021).

Randomized search explores only a subset of the specified hyperparameter values based on a random uniform distribution (Bisong and Bisong 2019). This approach can be more efficient than grid search when the search space is of high dimension and an exploration of all combinations is not necessary (Valarmathi and Sheela 2021).

5.2.4 Machine learning models

Two ML models was developed in this study, namely a Random Forest model and a Multilayer Perceptron model. Both models has been identified in the literature as favorable models for lead time prediction tasks, see section 3.4.

Random Forest model

The random forest model was developed with Sklearn ensembling method RandomForestRegressor. The method fits a number of classifying decision trees to different dataset subsamples. Averaging is used to increase the accuracy and reduce overfitting (Pedregosa, Varoquaux et al. 2011).

The hyperparameters used in the hyperparametertuning is shown in Table 11 along with a description of the hyperparameters.

Hyperparameters for RF	
Hyperparameter	**Hyperparameter description**
N_estimators	The number of decision trees in the forest
Min_samples_split	The minimum samples required to split an internal node
Min_samples_leaf	The minimum samples required to be at a leaf node
Max_features	The maximum number of features to consider for a split
Max_depth	The maximum depth of each tree in the forest
Bootstrap	If bootstrap samples are considered when building trees (If False the whole dataset is used to build each tree)

Table 11: Hyperparameters for RF

Multilayer Perceptron model

MLP is often mentioned as a favorable ANN architecture for prediction tasks (Ahmarofi, Ramli et al. 2017). The MLP model applied in this study was developed using MLPRegressor from Sklearn's neural network module. The MLPRegressor apply backpropagation for training and has no activation function in the output layer.

The hyperparameters used in the hyperparametertuning is shown in Table 12 along with a description of the hyperparameters.

Hyperparameters for MLP	
Hyperparameter	**Hyperparameter description**
activation	Function that determines whether or not a neuron should fire
solver	Solver used for weight optimization
alpha	Regularization parameter that can help prevent overfitting
hidden_layer_sizes	The amount of neurons in each hidden layer
max_iter	Maximum number of iterations

Table 12: Hyperparameters for MLP

6 Results

The findings from this master thesis are presented in this section. This includes results from the company interviews on PPC lead time applications for the fenestration industry, data quality assessment and machine learning experiments.

6.1 Results of PPC application areas for lead time prediction from company interviews

Interviews were conducted with Gilje, Nordic Door and MagnorVinduet in order to map the relevancy of using LTP for the different application areas identified in section 3.2.2. The results from the interviews are summarized in Table 13.

Gilje

The responses from Gilje were based on fixed casement windows, which are their most frequently produced window. In their production planning, Gilje has expressed an interest in using a LTP model. Since they have installed an RFID system, the company has valuable data that can be used for a number of the application areas mentioned in the framework.

Gilje's capacity planning is today based on the amount of glass panes in a window. This is a rough measurement which only considers one order characteristic, and does not take into account happenings at the shop floor or system load. Gilje recognized the potential for employing lead time to replace their current capacity planning strategy, in order to improve the accuracy of their planning by including several dynamic variables. The relevancy of integrating LTP in their due date setting was set to low, because the customer is essentially the one who requires a specific delivery date. Gilje determines whether they have available capacity on that particular day, and if not, negotiations with the customer will take place. Gilje did however note that they typically have enough capacity to accept all orders. The company showed a moderate interest in employing LTP for their production scheduling. They follow a FIFO principle, where lead time is of little interest.

Production analysis was seen as highly relevant for Gilje. The company's RFID system provides them data that can be used to predict the lead time of future orders. However, as of today, they use the RFID data solely for real-time tracking at the shop floor. Gilje saw a great potential of employing LTP to analyse various aspects of their production planning, such as determining the best product mix for optimizing the production flow. Additionally, they recognized the potential of employing LTP to dynamically determine the ideal WIP and buffer levels, particularly for the buffer between the priming and top-coat area. Gilje expressed a moderate interest for production monitoring. Their five-day lead time buffer reduces the risks and uncertainties related to lead times and and can shield the delivery precision from production delays. A comprehensive production monitoring is therefore not of highest interest. On the other hand, a strong interest was shown to reduce these buffer days through increased planning precision, which can be achieved with LTP.

Nordic Door

Nordic Door's response was based on their most popular product, the standard office doors. They discussed the value of lead time and how it can be used as a key performance indicator, adding that the company could be able to gain an overview of their performance if they could predict precise lead times. Several of the application areas from the framework were considered as highly relevant for the company.

When it came to capacity planning, the organization exhibited a strong interest in employing lead time to determine the capacity. As mentioned in section 2.2, today they use the amount of doors as a capacity measure, and neglect shop floor happenings and the system load, similarly to Gilje. However, they recognized the possibility to estimate capacity based on the predicted lead time of a door rather than the number of doors. Due to the fact that the customer usually sets the due date, there was only a moderate level of interest in integrating LTP with the company's due date setting. Nevertheless, they noted that incorporating lead time while estimating the delivery date could increase the delivery accuracy if a negotiation about the due date arose. In Nordic Door's production scheduling LTP was, similarly as due date setting, considered to be of moderate interest. They stated that lead times aren't particularly important in the daily production schedule that they use today.

Nordic Door saw a lot of potential for both tasks within the production control, namely production analysis and production monitoring. But as of right now, none of the tasks can employ lead time data since the company doesn't collect any data along the production line. However, they acknowledged that if they had sufficient data, it would be extremely beneficial to use LTP for both production analysis and monitoring. The company would be able to understand the production better by undertaking a production analysis, and predict the influence of future changes. They added that monitoring with remaining LTP could be a useful strategy for locating the source of production delays in real time, which could then be used to reschedule and improve the ongoing production.

MagnorVinduet

In general, MagnorVinduet showed a moderate to low interest in using LTP model for the different application areas. Regarding capacity planning, the relevancy was set to moderate. They stated that LTP could be relevant for the parts of the production that had a linear layout, i.e. LTP for each work station. A LTP for the entire production, however, would not be of relevance because of their high product variation, and they further claimed that the variations are too unique for a ML model to detect any patterns. They added that they were quite satisfied with the capacity planning approach they had today, which was based on cost and time estimates. The employment of LTP for due date setting had a low level of significance. The company does always assume the lead time to be around one week, due to the application of a weekly production schedule. A more precise estimate of the lead time would therefore not provide any value to their current due date setting. However, they mentioned that the manufacturing is largely dependent on supplier delivery times, and that these times set boundaries for when the company can promise a delivery date. A prediction of the supplier lead time may therefore be of significant relevance to the company. LTP to increase production scheduling was also set to low relevance. The company asserted that there were numerous elements that affect lead times, and, as for capacity planning, their production exhibits no distinct patterns that a potential ML model could identify and subsequently use for scheduling purposes.

MagnorVinduet placed both production control tasks under low relevancy. As mentioned above, the company believes there are too many factors influencing the lead time for a machine to accurately estimate it. Therefore, they argue that using LTP for production analysis is not particularly realistic. They think it will take more time to create and work on digitizing this task than doing it manually, as they do now. MagnorVinduet showed some interest in using historical data to verify the time estimates provided from their data system. These time estimates, however, take into account more than just the lead time as it is described in section 2.2. Additionally, the company lacks historical data on the actual time an order spends in the production, thus in order to do a comparison, they must first put in place a system that can collect this data. With regard to the final task, production monitoring, MagnorVinduet once again believed that this would not be pertinent for a company as specialized as theirs. As mentioned in section 2.2, the company had previously implemented an RFID system that provided them the necessary data for this task. However, none of the employees recognized the possibility for utilizing this information in any aspect of the production planning. The company further indicated that employing ML to estimate remaining lead time will not enhance productivity beyond what is now accomplished by workers.

The interviews revealed a difference between the companies' responses about the use of LTP in the five application areas. Nordic Door and Gilje, both saw potential in applying LTP for capacity planning. The planning accuracy may be increased by including LTP in their capacity approach, since the method will consider a number of factors, compared to todays' one factor strategy. MagnorVinduet employed several factors in their current approach, and expressed only a moderate interest for capacity planning. All three companies showed a moderate to low interest for integration of LTP in due date setting. Due dates are typically specified by the customer, therefore precise lead time predictions are not needed. Gilje and Nordic Door both have a daily scheduling approach and placed production scheduling under moderate relevance, while MagnorVinduet has a weekly approach and placed it under low relevancy. None of the companies have employed a scheduling approach based on lead times. Consequently, they did not perceive a clear necessity for the application of LTP in this field. Both Gilje and Nordic Door recognized the value of using LTP for production analysis in order to gain an overview and enhance their production performance. For production monitoring, Nordic Door was highly interested while Gilje showed a moderate interest due to their five-day lead time buffer. MagnorVinduet showed a low interest for both production analysis and monitoring, due to the numerous factors affecting lead times and the lack of distinct patterns in their production. From the answers provided by the three interviewed companies, a summary of the LTP application areas is presented in Table 13. The authors asked the companies if there were any further application areas where lead time could be of relevance, but no additional areas were found during this conversation.

Application Area	**Company**	**Relevance**		
		Low	**Moderate**	**High**
Capacity Planning	Gilje			✓
	Nordic Door			✓
	MagnorVinduet		✓	
Due date setting	Gilje	✓		
	Nordic Door		✓	
	MagnorVinduet	✓		
Production Scheduling	Gilje		✓	
	Nordic Door		✓	
	MagnorVinduet	✓		
Production analysis	Gilje			✓
	Nordic Door			✓
	MagnorVinduet	✓		
Production monitoring	Gilje		✓	
	Nordic Door			✓
	MagnorVinduet	✓		

Table 13: Summary of interview response for PPC application areas for LTP

6.2 Results from data quality assessment

The results from the data quality assessment were achieved through quantitative measures of the data utilized for the ML task, additional interpretation was achieved through communication with Gilje. The assessment was conducted in order to evaluate the completeness, consistency, accuracy and believability of the data.

Data from four independent data sources, presented in section 5.1.1, were used for the assessment, however dataset 1, the RFID data, has received more weight due to it's importance for LTP. The data dimensions included in the evaluation were derived from the literature, and focus on the dimensions highlighted for big data and sensor implementation: completeness, consistency, believability and accuracy. Several data quality issues may belong in more than one dimension, but for the sake of clarity, they are presented under the dimension that the authors deem to be most appropriate. The simple ratio form for completeness (Comp), consistency (Cons) and accuracy (Acc) presented in section 3.5 are used for the assessment:

$$Comp = \frac{Nr.\ of\ missing\ values}{total\ nr.\ of\ values\ in\ the\ dataset}$$

$$Cons = \frac{Nr\ of\ values\ that\ respect\ the\ constratint}{total\ nr.\ of\ values\ in\ the\ dataset}$$

$$Acc = \frac{Nr.\ of\ correct\ values}{total\ nr.\ of\ values\ in\ the\ dataset}$$

Through the conducted case study several indicators for data quality issues were mentioned from Gilje regarding the RFID data. In section 4.4, Gilje's utilization of their RFID system was presented. Gilje communicated that the automation of the packing machines through RFID data had medium satisfaction. This was mainly because it appeared faults in both RFID tags and the tags falling off from the components, resulting in stops in the machines and need for manual assistance. MagnorVinduet stated the same problems with their previous RFID system, presented in section 2.2. Real-time update of reordered components using RFID achieved high satisfaction at Gilje, but they do not have any quantitative measurements of the actual reliability of this system. According to Gilje, it occasionally happens for the RFID tag on the component to get lost, or the operators forget to register the components out of the system at the end of production. Additionally, they noted that, in rare cases, an RFID tag may be affixed to the incorrect component or contain inaccurate information. Gilje's data systems showed that the level of connectivity was fairly low in the manufacture. The RFID data was not integrated with Gilje's planning system or analyzed a posterior. This results in low level of knowledge and utilization of the produced data.

Completeness

Determining whether there exists sufficient and suitable amount of data for the task (Cichy and Rass 2019).

The evaluation of completeness was conducted on dataset 1 and 2, since these were the datasets used for the intended task, ML-based LTP. The sufficient amount of data for a ML model is task dependent and vary depending on the models complexity. A thumb rule is to have sufficient amount of data in order to represent all categories.

In order to address the datasets ability to represent different categories and variations, an in-depth analysis was conducted on several features. Dataset 1 contained 287 014 rows and was captured over a period of around six months. The data assessment revealed that some variables, such as *skråtapping* and *type of wood*, in this time period had a constant value across the entire dataset. The height and width of a component could vary between 25 cm to 7000 cm, and the customer could choose between over 2000 colors, as presented in section 4. Dataset 1 and 2 contained 168 distinct colors, 623 different widths, and 691 different heights. The distribution of the heights and widths are presented in Figure 28. The unique values are not distributed equally, as can be observed in the figure, and the model will be impacted by some values that occur more frequently and hence receive a larger weight during model training. The dataset only achieved 8,4% completeness from measuring the completeness in terms of colors.

$$Comp = \frac{2000 - 1832}{2000} = 0.084$$

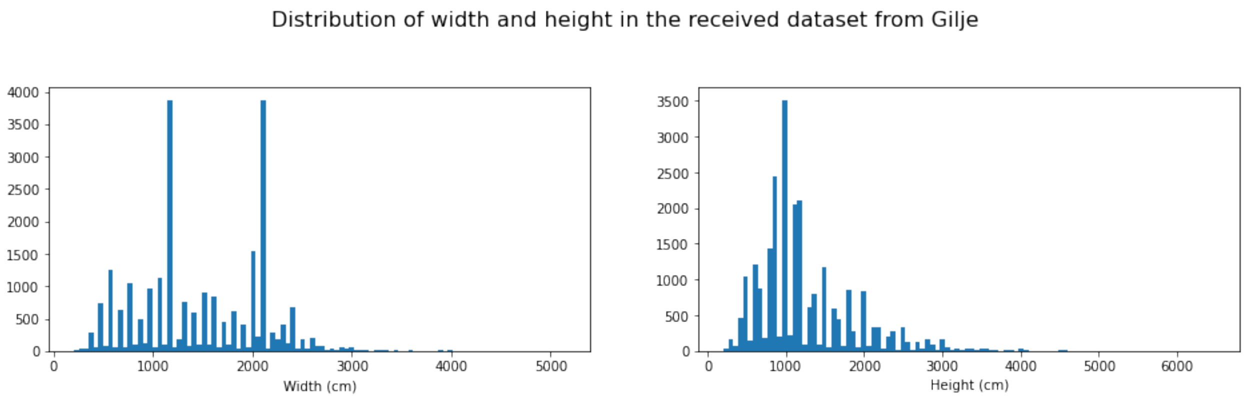


Figure 28: Distribution of width and height in dataset 1 and 2.

The number of RFID scanning points each component went through was another quantifiable property in dataset 1. According to the authors business understanding, all components that were not manually labored or reordered should have passed the shop floor scanning points presented in Figure 14 in section 4.4. Dataset 1 had a total of 271 941 data records after the manual labored and reordered components were removed. A total of 108 942 data records were found to be missing one or more shop floor scanning points. The result from the below equation show that only 60% of the components successfully navigate through all shop floor scanning points:

$$Comp = \frac{271941 - 108942}{271941} = 0.599$$

Consistency

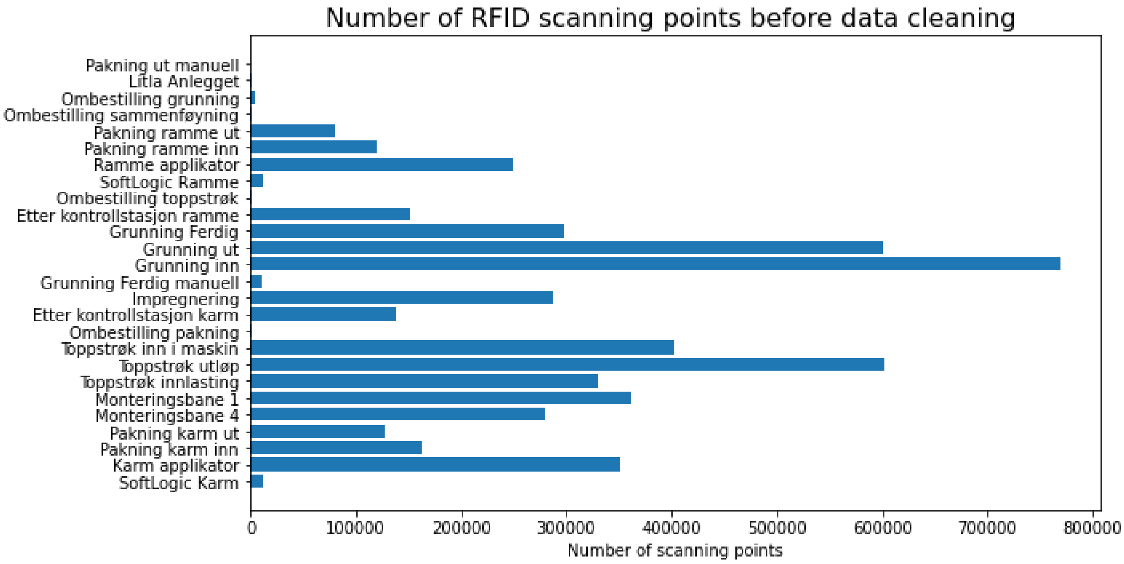
The extent to which the data is consistent with earlier data and data constraints (Taleb, El Kassabi et al. 2016).

The lack of completeness of shop floor scanning points in dataset 1 affects the consistency as well. The interquartile range constraint, presented in section 5.1.2, was used to measure the consistency of the calculated lead time feature. Several lead times in the outer scope of the mean were found from the outlier detection carried out with IQR, particularly in the upper quartile which resulted in very long lead times. A max lead time of over 120 000 seconds was detected before IQR was applied, with a mean of 25 000 seconds. After removal of the data records where it appeared incomplete shop floor scannings, 822 data records were found to not satisfy the IQR constrain. The total number of data records in the dataset was then given as: (*total number of records* − *records with incosistent shop floor scanning points*) = (271941 − 108942) = 162999. The consistency can be determined by the below formula:

$$Cons = \frac{162999 - 822}{162999} = 0,995$$

Only 0.5 % of the dataset is not consistent with the IQR constraint. The dataset may still contain faulty data records that the IQR has not been able to identify, because there's a probability that the dataset contains additional measurement mistakes which are not yet detected and could have an impact on the calculated lead times.

Figure 29 depicts the number of shop floor scanning points for each RFID scanner before and after the removal of the components that did not pass all of the scanning points, referred to in the figure as data cleaning. After data cleaning, each scanning point's consistency increased, but as can be seen in the figure there still occurred some inconsistencies in the data. For instance, both top coat scanning points, *Toppstrøk inn i maskin* and *Toppstrøk innlasting*, should have been passed the same number of times for all components and should therefore have been consistent, but they differed by about 50 000 scanning points. The same applies for *Grunning ferdig*, which



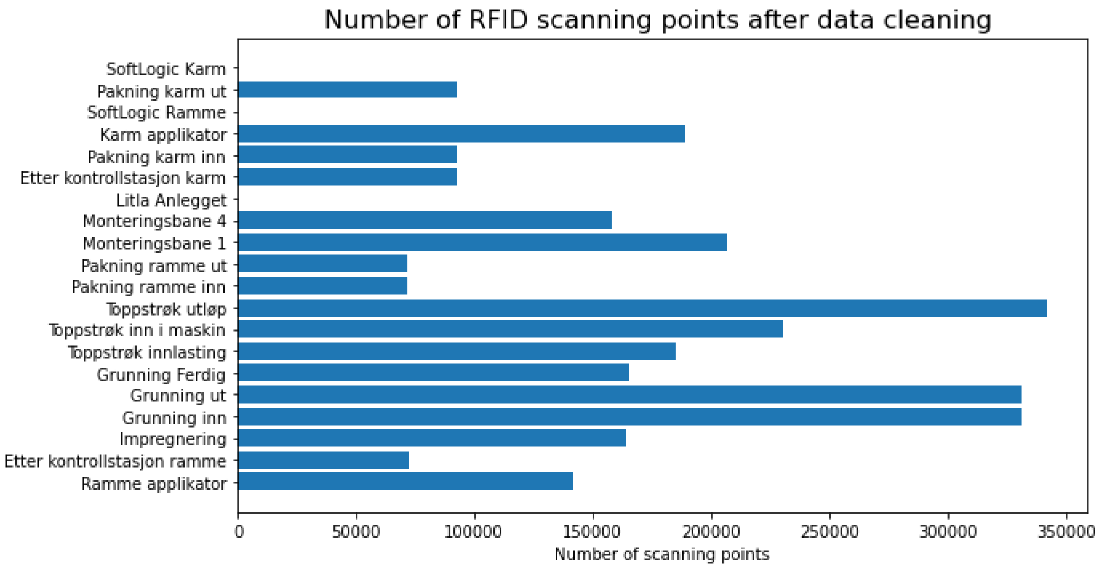


Figure 29: Number of RFID shop floor scanning points for all data records before and after data cleaning.

was expected to be consistent with the top coating scanning points. The top coat out scanning point, *Toppstrøk utløp*, priming in, *Grunning inn*, and priming out, *Grunning ut*, were expected to have several instances as shown in the figure, since most components pass these scanning points twice. A deviation between these scanning points are reasonable since it is customer specific if the priming is applied once or twice, or even at all.

The number of scanning points for each component was also taken into consideration while assessing the RFID system's consistency. The average number of scanning points per component was nineteen. Under normal production a component has sixteen scanning points, including the scanning points during assembly. In Figure 30, the distribution of the number of scanning points for each component before data cleaning is presented. In the figure, several inconsistent scanning points can be observed, where the maximum number of scanning points a component had was 27 494.

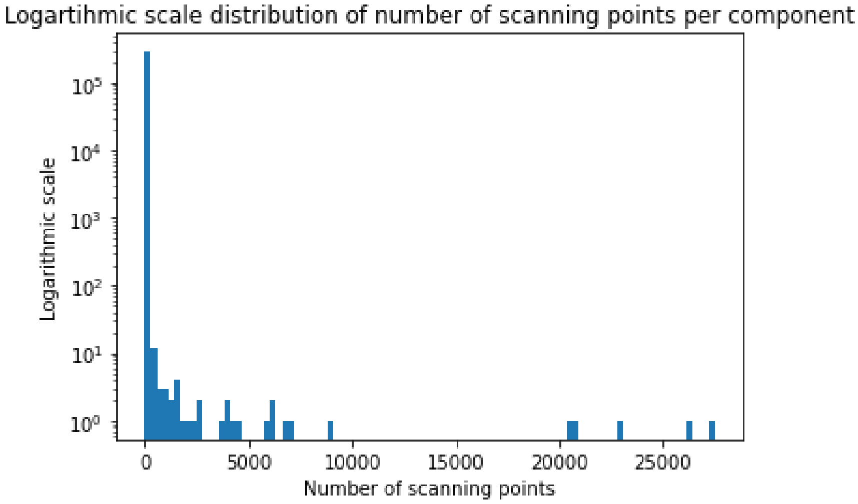


Figure 30: Logarithmic distribution of number of scanning points per component

Believability

Specifies whether a data source generates true, reliable and credible data (Juddoo 2015).

The believability of the RFID system was called into question by the aforementioned incorrect scannings detected in dataset 1. In order to evaluate the plausibility between the datasets, as well as the believability of the data sources, similar attributes in the datasets were compared.

The time stamp of the last shop floor scanning point in dataset 1, and the actual delivery date of an order in dataset 2 serve as a quantifiable consistency constraint between the datasets. 9623 data records were detected with actual delivery date before the last shop floor time stamp. A further examination showed that 5680 of these data records had the actual delivery date before the first shop floor scanning as well. The remaining 3943 data records had the actual delivery date between the start and end of the shop floor scanning points. The consistency/believability can be measured as:

$$Cons = \frac{276912 - 9623}{276912} = 0,965$$

3,4% of the data records indicated discrepancy between the two features. Gilje has stated that the delivery date may be changed in consultation with the customer without updating the data records in dataset 2. Changes in the delivery date are frequently made at the request of the customer, or as a result of delayed production. An analysis of the 9623 data records revealed an average temporal difference between the two features of 23 days. 25% of the data records had a shop floor time stamp more than 44 days after the actual delivery date, with a maximum value of 152 days.

Accuracy

The evaluation of whether the data is correct, reliable and certified (Cichy and Rass 2019).

In this data quality assessment the accuracy was viewed as the total amount of detected errors in the input data for the ML model. After the performed data preparation, presented in section 5.1.2, the dataset contained a total of 128 888 rows, which was reduced from the original 276 912 records. The accuracy of the dataset is presented in the formula below:

$$Acc = \frac{128888}{276912} = 0,465$$

This results in 47% accuracy for the final input dataset for the ML model, which contains information from both dataset 1 and 2. As mentioned, there still exists inconsistency in the number of shop floor scanning points, which affects the calculated measures and among them the target variable, the lead time, for the ML model.

In conclusion, a number of flawed data records and inconsistent values were found during the data quality assessment. The RFID scanning points in dataset 1 account for the vast majority of defects. Additionally, discrepancies in values across datasets 1 and 2 were found, which weakened the credibility of the sources.

6.3 Results from machine learning experiment

The purpose of the the ML experiments was to examine the performance of different ML models using the evaluation metrics described in section 5.2.1. In order to determine which model performed the best, the models were trained using various hyperparameters and datasets.

6.3.1 Experiment 1: Model performance with and without system status data

Several articles highlights the importance of including system status when predicting lead times. To investigate the impact of system status, the authors generated two datasets. The first dataset, *ML_dataset_sys*, included both product characteristics and system status information such as wip and the shop floor load, whereas the second dataset, *ML_dataset*, solely included product characteristics. The performance of a RF model and a MLP model were evaluated and compared with a baseline linear regression model. In experiment 1, the models were trained without hyperparameter tuning, and default hyperparameters were used for all models.

Linear Regression				
Dataset	**R^2**	**MAE**	**MSE**	**RMSE**
ML_dataset	0.0327	0.8012	0.9724	0.9861
ML_dataset_sys	0.0316	0.8100	0.9738	0.9868

Multilayer Perceptron				
Dataset	**R^2**	**MAE**	**MSE**	**RMSE**
ML_dataset	0.2104	0.6902	0.7937	0.8909
ML_dataset_sys	0.5703	0.4635	0.4321	0.6573

Random Forest				
Dataset	**R^2**	**MAE**	**MSE**	**RMSE**
ML_dataset	0.3083	0.6040	0.6953	0.8338
ML_dataset_sys	0.6743	0.3566	0.3275	0.5723

Based on the presented evaluation metrics, the findings from the experiment showed that using *ML_dataset_sys* improved the models performance for both the RF model and the MLP model. This suggests that an inclusion of system status features enhances the models ability to make accurate predictions. For both MLP and RF, the R^2 evaluation metric increased with the inclusion of data from the system, while there was little difference in the outcome for the LR model. MAE, MSE and RMSE showed a decrease when including the system status for the RF model and the MLP model. Again, LR did not drastically change.

By comparing the ML models' performance, both RF and MLP outperformed LR for all evaluation metrics. These models demonstrated higher predictive accuracy and were able to capture more complex relationships within the data. As a result, they are considered to be better suited for modeling the given problem.

Overall, the findings from the first experiment highlights the importance of incorporating system status data, and emphasize the superiority of RF and MLP models over LR for achieving better ML results for LTP.

Feature importance

Further, the feature importance from RF was evaluated, presented in Table 14. The features with an importance rate of 0.05 or higher are presented in the table. These results also showed that the system status features had the most significant impact on the models. The features and their importance are also visualized in Figure 31, where the features with 0.0 importance are excluded.

Feature	RF-importance
count_comp_level	0.22
series_count	0.21
count_buffer_level	0.19
length	0.08
prods_belegg	0.05
ant_karmkomp	0.05
ant_rammekomp	0.05

Table 14: Feature importance from the most influential variables from the RF model

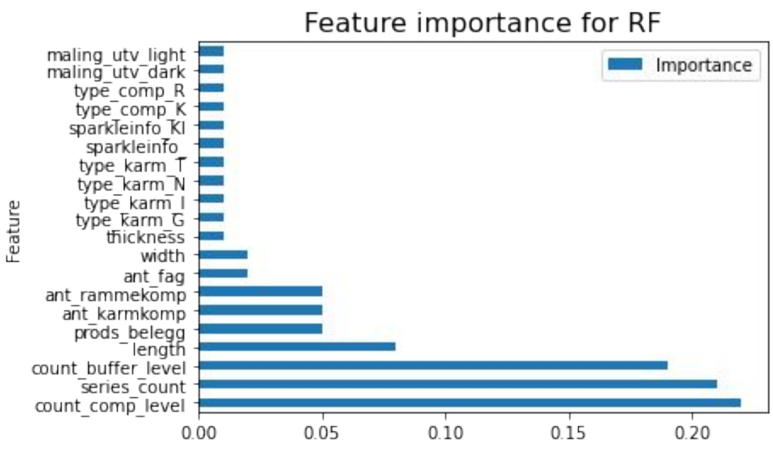


Figure 31: Feature importance from the RF model.

6.3.2 Experiment 2: Hyperparameter tuning of random forest and multilayer perceptron

In experiment 2, hyperparameter tuning of the RF model and the MLP model was performed in order to identify the optimal model architecture. Both *randomized search* and *grid search* were used to find the optimal parameters. Randomized search was initially used to acquire a feel of the general range and distribution of hyperparameter values that provided a satisfactory result, because it is quicker than grid search. An exhaustive search using grid search was then performed on a narrowed down grid based on the results from the randomized search.

Several experiments were executed with both randomized search and grid search, and the best performed search are presented below.

Experiment 2.1: Hyperparameter tuning for RF

The best results from the randomized search and the grid search are presented in Table 15. The optimal parameters obtained from the randomized search were used to narrow the search range for the grid search, and the best performing parameters from both experiments are highlighted in bold.

All randomized search experiments resulted in min_samples_leaf = 1, max_features = sqrt, and bootstrap = False. These values were therefore considered constant in the grid search experiment.

Hyperparameter tuning RF		
Hyperparameter	**RandomizedSearch**	**GridSearch**
n_estimators	[200, **400**, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000]	[300, **500**, 700]
max_features	["auto", **"sqrt"**, 1, 3]	[**"sqrt"**]
max_depth	[10, 29, 48, 67, 86, **105**, 124, 143, 162, 181, 200, None]	[105, **110**, 130]
min_samples_split	[2, 5, **10**]	[**5**, 12]
min_samples_leaf	[**1**, 2, 4]	[**1**]
bootstrap	[True, **False**]	[**False**]

Table 15: Hyperparameter tuning of RF

The parameter grid for the best model consisted of the hyperparameters in bold from the grid search in Table 15. The number of trees in the forest, n_estimators, of the final model was *500*, with a max depth of *110*. The minimum sample split was set to *5*, and the minimum sample leaf was *1*. It was not applied bootstrap and max_features was set to *sqrt*, which means that the size of the subset was sqrt(n_features), where n_features was the number of features in the dataset.

The distribution of the actual and predicted lead times detected by the RF model is shown in Figure 32. The RF model was able to capture some properties of the actual distribution, this corresponds to the statics presented in Table 16. 25% of the predicted values had lead times below 18 000 seconds, as opposed to about 15 500 seconds for the actual values, which is indicated by a smaller apex in the distribution. The same holds true for 75% of the predicted values, which had lead times that are less than 31 500 seconds, and 33 700 seconds for the actual lead times. This results in a lower peak at the end of the distribution for the predicted lead times in Figure 32. The model captured roughly the same mean, minimum and maximum value as the actual lead time. The two distributions show some variation in the standard deviation.

Static	**Predicted**	**Actual**
mean	25 096 sec	25 139 sec
std	10 551 sec	12 600sec
min	2 837 sec	2 278 sec
max	60 779 sec	61 557 sec
25%	18 053 sec	15 771 sec
75%	31 487 sec	33 706 sec

Table 16: Statistical values for predicted and actual lead times for the RF model

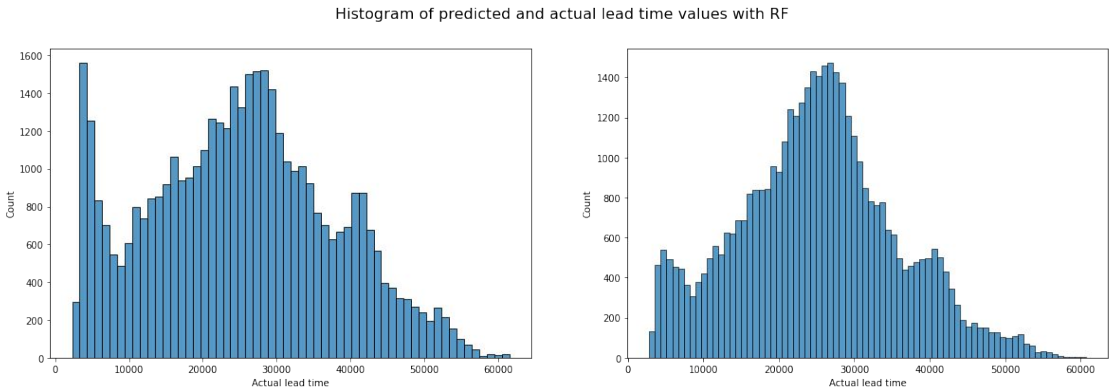


Figure 32: Histogram of the actual lead times and the predicted lead times from the RF model

Experiment 2.2: Hyperparameter tuning for MLP

Table 17 displays the hyperparameters that were used during the randomized search and grid search for MLP. The parameter values that yielded the best result are highlighted in bold. The best parameters from randomized search where utilized to narrow the search range for grid search. Value options were only provided for two hyperparameters in the grid search; hidden_layer_sizes and max_iter, since the other hyperparameters had a constant value in all the randomized search experiments.

Hyperparameter tuning MLP		
Hyperparameter	**RandomizedSearch**	**GridSearch**
activation	[**“tanh”**, “relu”, “logistic”]	[**“tanh”**]
solver	[**“adam”**, “lbfgs”]n	[**“adam”**]
alpha	[**0.0001**, 0.05]	[**0.0001**]
hidden_layer_sizes	(150, 100, 50), **(130, 100, 50)**, (120, 80, 40), (100, 50, 30), (120, 80)	**(140,110,60)**, (130, 100, 50), (120,80,40)
max_iter	[100, **500**, 700]	[**400**, 500, 600]

Table 17: Hyperparameter tuning of MLP

Based on the results displayed in Table 17, the authors decided to continue with *tanh* as activation function, *adam* as the solver for weight optimization, an alpha value equal to *0.0001*, *three* hidden layers with the sizes *140, 110 and 60* were employed and max iteration was set to *400*.

The MLP model was able to capture some patterns, as shown by Figure 33, however there was a significant discrepancy between the predicted values and the actual values. In Figure 33, it can be seen that the model was unable to detect the graph's first apex. Some of the predicted values were negative, which indicates poor model performance considering that lead times can not be negative. This is also demonstrated in Table 18, where the minimum of the predicted values were equal to -17 903 seconds, which is not a realistic lead time. Table 18 showed that the maximum lead time was longer for the predicted values than for the actual values, with a deviation of 14 185 seconds. Deviations between the actual and predicted values for the 25th percentile was also found, as 25 % of the predicted values dropped below 17 634 seconds, compared to the actual values, where 25 % dropped below 15 771 seconds. The standard deviation was higher for the actual values than for the predicted values, which indicates that the model has not been able to capture all variability. As for mean and the 75th percentile, both are higher for the actual values.

Static	**Predicted**	**Actual**
mean	25 139 sec	24 939 sec
std	12 600 sec	10 931 sec
min	2278 sec	-17 904 sec
max	61 557 sec	75 742 sec
25%	15 771 sec	17 634 sec
75%	33 706 sec	31 868 sec

Table 18: Statistical values for predicted and actual lead times for the MLP model

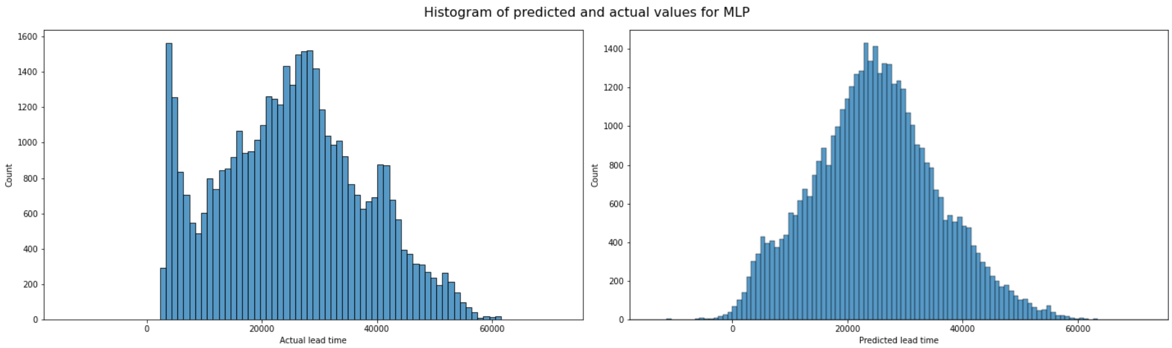


Figure 33: Histogram of the actual lead times and the predicted lead times from the MLP model

Performance metrics for RF and MLP

Table 19 displays the results from the evaluation metrics after the RF model and the MLP model were trained with the obtained parameters from the hyperparameter tuning. The results demonstrated that the RF model outperformed the MLP model. The R^2 evaluation metric indicated that the RF model was able to capture more of the variance compared to the MLP model. Deviations between actual values and predicted values were also lower for the RF model, as seen by the four other evaluation metrics.

Performance metrics for ANN and RF				
Dataset	$\mathbf{R}^2$	**MAE**	**MSE**	**RMSE**
MLP	0.6132	0.4021	0.3681	0.6067
RF	0.7110	0.2916	0.2906	0.5391

Table 19: Performance metrics for ANN and RF

In summary, the ML experiments proved that incorporation of system status variables increase the model performance. In addition, the LR model was outperformed by both RF and MLP . Lastly, hyperparameter tuning detected the best architecture for RF and MLP, where RF showed superiority over MLP for all evaluation metrics.

7 Discussion

This section addresses and discuss the three research questions of the thesis. The discussions are based on the conducted literature review in section 3, the case study in section 4, the applied data analysis in section 5, and the obtained results in section 6.

7.1 RQ1: In what areas of production planning and control can lead time prediction be applied in MTO fenestration companies?

Lead times have proven to be a key performance indicator in production planning and control, and with increased customization and order variety in MTO firms, LTP has become a more challenging task. Several authors have investigated approaches for LTP in the field of MTO companies. However, few articles investigate the actual applicability of LTP in PPC tasks, and no research was obtained in the fenestration industry. To address RQ1 a literature review was conducted in section 3.2.2, which identified five application areas where LTP can be applied within PPC; *capacity planning*, *due date setting*, *production scheduling*, *production analysis*, and *production monitoring.* The transferability of these PPC tasks to the MTO fenestration industry was discussed with three companies within this sector, and the results were presented in section 6.1. In this section, the findings from the interviews and the literature review are used to discuss and answer RQ1.

Regarding the interest in using LTP for different PPC activities, there were notable discrepancies between the interviewed companies. Nordic Door and Gilje both acknowledged the value of lead time as a key performance indicator, and responses from the companies indicated several potential application areas for LTP within their PPC. MagnorVinduet on the other hand, exhibited only a moderate to low interest for the employment of LTP within the different application areas.

Capacity planning

LTP integration for capacity planning was highly desired by Gilje and Nordic Door. Lead time has been cited in the literature as being highly important for capacity planning, and it can assist production planners in acquiring a more realistic capacity plan. This could also apply for the two companies. Both companies recognized limitations regarding their current capacity approach, which is based on the number of doors and the number of glass panes in a window. They might potentially develop a more precise capacity strategy by replacing their existing capacity planning method to one based on LTP. Compared to the one-parameter based approach currently used by the companies, LTP can consider shop floor load, multiple order parameters, and dynamic changes at the shop floor, which is likely to better reflect real-world capacity.

MagnorVinduet did only show a moderate interest for the use of LTP in their capacity planning. Today, they plan capacity based on rough price- and time estimates. MagnorVinduet's capacity planning approach incorporate several dynamic parameters and order specific factors, compared to

Nordic Door and Gilje, which can be a reason for lower interest. On the other hand, MagnorVinduet does not take system status into account in today's planning method, and they do not have a measure of the actual time consumption of their orders. Therefore integrating LTP in their capacity planning method could increase the planning accuracy. MagnorVinduet also expressed that LTP was primarily relevant for manufacturing segments with linear layout, but since the overall layout is functional with many routing variations an LTP model was not of high interest. In addition, the manufacture contains a lot of manual labour, where it is hard to retrieve precise data records. Prior research have proven that ML-based LTP can overcome challenges related to advanced manufacturing layout, e.g. in the research of Alenezi, Moses et al. (2008) a multi-resource and multi-product system was investigated where ML-based LTP showed promising results. Also challenges related to poor data records due to manual labor have been resolved in earlier studies, as seen in the textile and garment industry (Cao and Ji 2021; Susanto, Tanaya et al. 2012). On the other hand, most research regard simple flow shops and semi-conductor environments.

MagnorVinduet uses a self-developed Excel system for their data management, while both Gilje and Nordic Door have implemented an ERP system. The ERP system can be utilized for integrating various production elements, and it may offer companies more sophisticated software capabilities and built-in features that can give them access to vital information and data about the production which can be used for LTP. Excel systems can also be effective tools, but frequently, the functionality must be set up by the company itself, which can take a lot of time and involve a lot more manual labor. This shows a higher level of digitization at Gilje and Nordic Door. The absence of an integrated system, such as an ERP system, might make it harder for MagnorVinduet to collect the information required for reliable LTP. As a result, they may perceive a lower interest in integrating LTP into their capacity planning.

The exploratory data analysis of Gilje in section 4.5 showed a great deviation between the actual and the planned delivery date, as well as the actual and planned amount of glass points. These deviations may be due to data not being updated when a customer make changes to the delivery date. However, it may also be due to the production capacity being exceeded, which can lead to production delays and deviations from the original plan. Additionally, Nordic Door acknowledged that they often suffer from production delay's due to capacity limitations. MagnorVinduet also noted that practicing overtime work is occasionally necessary to deal with capacity surpassing. These findings may indicate an insufficient capacity planning. Kramer, C. Wagner et al. (2020) pointed out how LTP could be used to determine the capacity of the production system, and all three companies may improve the performance of their production planning by applying LTP for this task. The findings imply that LTP can be employed to optimize capacity planning within PPC for fenestration companies.

Due date setting

In terms of due date setting, Nordic Door notes a moderate level of interest in integrating LTP, whereas Gilje and MagnorVinduet reports a low level of interest. All companies determine the delivery date in accordance with the preferences of the customers, and if the company does not have capacity, negotiation may be required. They all stated that they usually have enough capacity to fulfill the customer-specified delivery date, since the customer often place an order weeks or months in advance of the delivery. Lead time is therefore not required for the determination of the delivery date, unless a negotiation arises. However in previous literature, due date setting has been of the most frequently mentioned application areas for LTP. R. Murphy, Newell et al. (2019) stated that accurate LTP may improve the delivery precision and hence increase customer satisfaction. In a number of the publications from the literature review, LTP techniques were utilized, and the outcomes were used to determine the order's due date (Gacek 2018; S. Hsu and Sha 2004; C. Wang and Jiang 2019). These methods differ from the due date setting methods applied by Gilje, Nordic Door and MagnorVinduet.

As mentioned, for all three companies, the customers frequently places an order weeks or months prior to the desired delivery date. This results in more time for production planning and coordination with the suppliers. Additionally, the companies have greater flexibility for modifications in the production process without it affecting the due date. Lead times may therefore be less critical for due date setting in this scenario. However if negotiation is required, incorporating LTP when determining the due date may help improve the delivery accuracy. This also applies for other companies where determination of the delivery date is dependent on negotiations with the customers, and where the company itself must develop a proposal for a delivery date. Findings from the interviews suggests that LTP may not be so relevant for companies within the MTO fenestration industry, since the due date is often set by the customer, and orders are frequently placed weeks or months in advance. However, LTP could potentially support negotiation processes that may arise when setting the due date.

Production scheduling

MagnorVinduet operates with a weekly production schedule, while Nordic Door has changed from weekly to daily production scheduling. Gilje operates with a daily schedule determined by a first-in first-out scheduling approach derived from the planned orders in the bar chart, presented in section 4.2. Gilje and Nordic Door both showed a moderate level of interest in employing LTP for their production scheduling. Neither of them uses lead time dependent methods for this task today. MagnorVinduet showed a low level of interest for the use of LTP in their scheduling approach. Similar to the two other companies, they adopt a scheduling strategy without considering lead time. They further claimed that LTP was irrelevant to their situation because there were no obvious patterns in their production. Without identifiable patterns, they might perceive the implementation of LTP as impractical or ineffective for their specific production context.

In the literature several articles highlight the importance of accurate LTP in order to optimize the production schedule (Bender and Ovtcharova 2021; Lim, Yusof et al. 2019; Lingitz, Gallina et al. 2018). In order for LTP to be relevant for the company's scheduling procedures it is likely that new and more advanced scheduling approaches are needed. Mori and Mahalec (2015) mentioned how lead times could be used to provide production schedules that are able to meet the production capacity and avoid delays and deviations. According to the findings from the interviews, mentioned in section 2.2, delays are occasionally a problem for the companies. The companies might need to re-prioritize the sequence of the orders in order to prevent delays. Dynamic scheduling and dispatching rules could potentially prevent or ease the task of re-prioritizing orders, hence reducing the amount of delayed orders. These delays and adjustments in the production could be costly for the company, and it could also affect customer relations. By implementing scheduling approaches which consider the anticipated lead times of the orders, the company may be able to obtain more accurate production schedules which in turn may decrease the risk of re-scheduling.

Production analysis

Through an examination of lead time data, the reason for production halts, deviations and delays could be identified, and through LTP, future halts, deviations, and delays could be anticipated (Pfeiffer, Gyulai et al. 2016). Nordic Door showed a strong interest in this application area, and suggested that LTP could provide an overview of the production performance. Gilje did also place descriptive analysis under high relevance. By analyzing lead times, they believed that they could obtain a better understanding of their production processes, optimize production flow, and determine the ideal product mix, which also could be anticipated for future orders with the use of LTP. Gilje might be able to fulfill more orders each day than they do now, by improving their production flow. Identification of factors that affect the production flow and the production mix, is also highlighted in the literature as important in order to increase the production performance (Choueiri, Sato et al. 2020; G. M. Lee and Gao 2021). Bottlenecks, which are emphasized as one of the primary causes for manufacturing delays, are another aspect that is mentioned by both Gilje and Nordic Door. The companies might be able to identify and eliminate the causes of bottlenecks by evaluating which features that are most influential for the LTP. The gained knowledge could potentially be used to take preventative actions. Both responses from the companies and findings from the literature highlights the potential for LTP to support the current and future production analysis. However, as of today Nordic Door lacks the necessary data and must thus put in place a data system that can capture lead time data before this application area can be applied.

Unlike the other two companies, MagnorVinduet placed descriptive analysis under low relevancy. The company expressed a preference for manual approaches over digitizing tasks like ML-based production analysis due to the complexity of developing and implementing such digital solutions. They have exhibited some interest in using historical data to verify the time estimates used in their planning. However, they lack information regarding the actual amount of time spent on historical orders to do so. MagnorVinduet is a smaller company than the two others, this might raise the barrier for installing more sophisticated and pricey technology. As mentioned above, they have also been sceptical about whether LTP is even realistic for a company with such specialized products. Nordic Door does also have a unique routing for all products, with a functional layout and long transportation routes. However they showed a strong interest for this task. Subjective opinions and research insight of the company representatives might be the cause of the different levels of interest. The representatives for Gilje and Nordic Door were both IT managers, while the representative for MagnorVinduet was the fabric coordinator.

Production monitoring

The last application area under discussion is production monitoring. This application field is mentioned in literature as a means of enhancing system effectiveness and customer relations (T. Chen 2009; Gyulai, Pfeiffer, Bergmann et al. 2018; Gyulai, Pfeiffer, Nick et al. 2018). As mentioned, production delays are a common occurrence, and dynamic re-prioritization of orders might decrease the number of delays. Production monitoring can work as a substitute or supplement to dynamic scheduling. By monitoring the production process and applying ML to predict the remaining lead times of the orders on the shop floor, companies may be able to adjust the job's priorities in a more efficient manner, which can prevent delivery delays of the orders. Nordic Door has placed production monitoring under high relevancy, because they suffer from a lot of delays. Using monitoring and remaining lead times to enhance the ongoing production is also highlighted in the literature (Gyulai, Pfeiffer, Bergmann et al. 2018; S. Huang, Guo et al. 2019). In contrast to Nordic Door, Gilje only expressed a moderate interest for production monitoring. The company has employed a five-day lead time buffer to mitigate production delays. This may indicate that production monitoring might not be as important when employing a lead time buffer in order to avoid delivery delays. On the other hand, through production monitoring Gilje might be able to reduce the amount of buffer days, which they have shown great interest in, by predicting the components actual termination times. Gilje added that production monitoring may be useful for identifying the cause of delays in real-time and make appropriate changes to ongoing production. The results indicate that LTP could be utilized for proactive management and real-time monitoring of manufacturing operations.

MagnorVinduet showed a low interest for production monitoring. They have previously implemented an RFID system which potentially could be used for monitoring. However, since the company discontinued this system, they are no longer able to perform this task with the data they have available. Once again, limitations in their production system could be a reason for the low interest. Though it's possible that the company can monitor production more effectively manually than the other two companies because it serves a smaller customer base in a more upscale market, negating the need for LTP in production monitoring.

Validity of the interviews

Three Norwegian fenestration businesses' were interviewed, which constitutes the findings shown in section 6.1. They all operate in an MTO manner and generate a wide range of customized orders. Different production layouts, ranging from linear to functional, are used by the companies. In comparison to the other two, Gilje has concentrated more on enhancing their manufacturing flexibility and efficiency, and they have a higher level of digitization. Older equipment, lengthy transportation routes, and a significant amount of manual labor are still present in Nordic Door's and MagnorVinduet's functional layout.

Gilje and Nordic Door were the companies with the most comparable responses. Compared to MagnorVinduet, they are both larger firms, which may boost their need for and ability to digitize their production. As previously indicated, the companies also have different production systems in place, with Gilje and Nordic Door being more technologically advanced. Handling the planning tasks manually may be simpler for MagnorVinduet since they are a smaller company with fewer employees and orders per day. MagnorVinduet had a poor implementation experience with the RFID technology, which could raise the bar for new digitization strategies. Additionally, MagnorVinduet operates with weekly production scheduling which might simplify their planning needs, compared to Gilje and Nordic Door's daily scheduling.

The three organizations represent some variety of the MTO fenestration industry, indicating a possibility of generalizing the results to businesses with similar operating characteristics. As mentioned, they produce highly customized windows and doors which is a common characteristic for fenestration companies. The companies cover a range of company sizes and customer groups. This diversity may enhance the validity and generalization of the findings, since the companies cover a wider variety of trends in the fenestration industry. However, the fact that all businesses are based in the same nation, may have an impact on the nature of the business and their products. Additionally, only one representative from Nordic Door and MagnorVinduet was interviewed, which might impact the results through subjective judgments and degree of expertise. In order to expand the findings to a broader scope of the MTO fenestration industry several companies and representatives in different nations should be included during the interview phase.

From the literature, five application areas have been identified where LTP could be used within PPC for MTO companies. The application areas were summarized in Table 3 in section 3.2.2, and include capacity planning, due date setting, production scheduling, production analysis and production monitoring. The framework was further discussed with three fenestration companies, namely Gilje, Nordic Door and MagnorVinduet. Gilje and Nordic Door recognized the potential of using LTP for various aspects of their production, including capacity planning, production analysis, and production monitoring. MagnorVinduet had in general a low interest of employing LTP in their production planning. The diverse perspectives of these companies underscore the importance of considering the companies' adoption of different PPC tools and production technologies when evaluating the potential benefits and challenges of implementing LTP. Factors such as the nature of the products, available data, current production processes, and prioritization of different operational aspects influence the perceived relevance and interest in employing LTP across different application areas. Regarding due date setting and production scheduling, the interest was significantly lower than the three other tasks. For many fenestration companies, due dates are set by the customer and are therefore not dependent on the lead time. However, LTP may be of relevance if negotiations occur. LTP has a varying degree of importance in production scheduling depending on the company's scheduling approach. Neither Gilje, Nordic Door or MagnorVinduet employ a lead time based scheduling approach. However, literature suggest that by employing LTP, dynamic scheduling approaches can be adapted and ease the phase of re-planning. This also applies for capacity planning, which both companies placed under high relevancy. Production systems within fenestration companies are often characterized by high complexity, due to highly customized products. As a result, production monitoring and analysis are crucial application areas that provide the company with a greater understanding of its production. By utilizing LTP within these areas, the company has the possibility to optimize the production and increase overall efficiency.

7.2 RQ 2: What can be the causes for data quality issues in data applied for lead time prediction?

Challenges related to data quality can have a substantial impact on ML-based LTP, as the accuracy, completeness, consistency and believability of the data are essential in order to achieve reliable and accurate outputs. As presented in section 3.3.1, regression models are trained on data that includes input/output pairs, and detects the patterns between the input and output variables. Poor data quality can lead to erroneous pattern detection, resulting in incorrect predictions. The main results from the data quality assessment showed that Gilje's RFID system has a number of data quality issues. In addition, the order data was affected by outdated records, and inconsistency between data sources were detected. In order to answer RQ 2 the detected data quality issues are discussed in the manner of causes detected in the literature and statements from Gilje. Additionally, the effect these data quality issues have on the ML models, as well as improvement suggestions are presented to highlight the impact on LTP.

The completeness of the data is essential for the ML models ability to capture the variations and trends in the window industry accurately. The RFID data was captured over a wide time period, six months, which the authors thought to be enough to represent the general characteristics and trends in the window industry in that period. However, the assessment showed that some variables had a constant value across the entire dataset, which indicated that the dataset did not include all variations. In addition, low completeness performance of window characteristics, such as color, width or height, could lead to incorrect predictions for new orders with other characteristics that were not captured by the training data. This suggests that, in order to detect new trends and changes in the market, it is crucial to update the training data for the ML model on a regular basis. The completeness of the color variable only achieved 8,4%, but this is likely a performance index that was significantly impacted by the trends in the market. Although there are 2000 hues available, it is likely that some of them are more popular among the customers, and this popularity may be determined by market trends. As a result, the authors did not consider this performance index to accurately represent the completeness of the data.

The completeness of the RFID scanning points was considered a more general and describing performance index since it is not affected by market trends, and should be able to capture all the necessary data records on the shop floor. The data quality assessment revealed that Gilje's RFID system failed to capture all necessary data related to the production process. Only 60% of the dataset were consistent with real-world expectations, and had records from all shop floor scanning points. These faults could lead to inaccurate lead time- and system status calculations, which are derived from the RFID data, resulting in incorrect data entries to the ML model. As detected through the IQR assessment, it is likely that several of the lead time values calculated by the authors were incorrect. The target variable for the model was the lead time. Therefore, incorrect lead time calculations would have a significant impact on the accuracy of the model. Additionally, the results revealed high inconsistency of shop floor scanning points for each component. The results also showed that the distribution of RFID scanning points still included inconsistent data after the incomplete data entries were removed, as presented in Figure 29. This could result in inconsistent values for comparable components that ought to have been equal. As a result, the ML models' ability to detect patterns between the variables could be reduced.

The deviations of shop floor scanning points could have several causes. They could occur from errors in the RFID scanners, faulty sensor node measurements, as highlighted by Alwan, Ciupala et al. (2022), such as multiple scans for a single component which causes redundancy in the data, sensor node malfunctions, calibration issues or poor sensor node quality, which can result in interrupted, noisy and missed scannings. Additionally, unexpected events may appear at the shop floor that the RFID system is unable to record. Gilje has stated that it appears that RFID tags falls off, leading to missing data, or operator's ignorance and negligence, like incorrect placement of the component which makes the tag invisible for the reader. The incomplete data records can also indicate lack of validation of the RFID system. Gilje's low utilization of the RFID data today might be a reason for this. Validation is time consuming and holds up resources, and therefore have to result in high profit for the company to prioritize this. Due to low validation, problems can happen during data transition, as a result of network or communication errors, leading to incorrect data entries as well. Gilje has not validated their RFID system, and measurement of the quality of data transition is therefore not feasible with the data at hand for this thesis. The reasons mentioned above aid in understanding the inaccurate data, but it's likely that the system has other flaws that are yet undetected.

On the other hand, shop floor scanning points' actual completeness might be greater than the findings. During data preparation, the authors employed a rigorous constraint that resulted in the elimination of all components that did not pass all eleven shop floor scanning points, illustrated in Figure 14 in section 4.4. This resulted in removal of every component that leaves and then returns to the manufacturing line, as described in section 5.1.2, involving manually processed and reordered components according to the authors. The RFID tracking of these components were accurate according to their routing, despite the fact that the data quality assessment flagged them as incomplete. Several undetected situations where a components routing was intended to leave and rejoin the production line could occur, which the authors have not detected through the case study. If there are several of these situations, the real-world completeness of the shop floor scanning points' will likely be higher than the findings.

Contradictory data sources were found from the data provided by Gilje, making it challenging to determine which data source that produced trustworthy data. The discrepancies between data sources and outdated data records decreased the believability of the data. Without accurate and trustworthy data the ML models prediction of future lead times are likely to be inaccurate as well, since the model learn erroneous relationship among the variables. 3,4% of the data, derived from dataset 1 and 2, had contradictory values, with the delivery times occurring before the registration of the last RFID scanning. Gilje has stated that this is likely due to obsolete delivery time data in dataset 2, as presented in section 4.5, which reduces the belivebaility of this dataset. Additionally, the production is unlikely to be delayed by several weeks, as the average temporal difference between the detected end of production in dataset 1 and actual delivery time in dataset 2 was 23 days. Which increases the likelihood of obsolete data records in dataset 2. But, for the 3943 components that had the first RFID scanning in advance of the actual delivery date, and the last RFID scanning after, the faults might occur from dataset 1. A possible cause that Gilje has mentioned, is that the operators needs to register the components out of the system at the end of the production. If they forget this, the components can be registered in the production for a long time. As well, the aforementioned reasons such as sensor node malfunctions, calibration issues, poor sensor node quality, and network or communication errors may also be an explanation.

The lack of consistent, complete and believable data results in low accuracy for the entire dataset. After data cleaning only 47% of the dataset remained. Even though several faulty records were removed there still existed faults in the dataset that were not detected, as was showed by evaluating the frequency of shop floor scanning points, presented in Figure 29. Training a ML model on faulty data result in low performance. This reduces the chance to successfully apply a ML model for decision-support or real-time event handling. Additionally, the reduction of the dataset also affects the ML models, since they in general perform better with a larger training set. The necessities of a large amount of data cleaning was proved to be essential with the data reviewed from the explored case study.

Data quality improvement

The literature on data quality assessment for big data and ML mainly highlights the possibility to detect and improve the data through data-driven improvement, namely data cleaning. Data cleaning techniques are widely applied through this study with the aim of reducing the impact of the detected poor data records. A drawback is the time consumption of data cleaning and the risk of removing important features, or missing the removal of faulty data records. The time aspect of data cleaning is specially important to consider if a predictive decision-support system was to be implemented in Gilje's planning system. As well, for ML purpose, the great reduction in usable data could be considered as a negative side effect, since ML models are highly affected by the data quantity, and more data could improve the accuracy of the model. Even though the data used in this study was greatly cleaned, there still existed faulty records in the final and cleaned dataset that were not detected by the authors.

In addition to data-driven improvement, process-driven improvement is highlighted in the literature for data quality improvement. In this thesis, the focus on improving and detecting when and where the faults occurred in the RFID system was specially important, in order to achieve accurate data records for the future. Several process-driven improvement steps could be adopted in order to mitigate these challenges. Implementing quality procedures and checks at the shop floor to ensure that the components are correctly placed according to the RFID scanner, and that the RFID tags are properly attached to the components could enhance the accuracy and completeness of the data. The system design also has improvement potential. By employing more RFID scanners in the production, the simplicity of detecting where the scanning faults occur could increase. Additionally, implementation of RFID scanners specifically for the buffer reduces the need for manual calculation of important system status variables, such as WIP, and will further provide important insight for Gilje. WIP levels are an important key performance indicator for many planning procedures as mentioned in section 3.2.2, and by providing accurate calculations for the WIP levels, the performance of LTP may be improved.

Company strategies can also be adopted in order to improve the data quality. Clear data quality standards can be implemented in the company, such as data accuracy, completeness, consistency, and believability, and ensure that these standards are consistently met. The company could carry out routine quality checks to make sure that certain measurements, such as the consistency of the shop floor scanning points, were above a specified threshold value. Implementing standard company procedures for updating manual changes in the data records through consistent use of IT solutions will increase the consistency and believability of the data. A better integration between data sources may also enhance these data quality dimensions. The company can implement data validation techniques to identify and correct data quality issues, such as missing data or incorrect data entries. Additionally, they can utilize data quality tools and software to automate the data cleaning and validation process, ensuring that data quality issues are quickly identified and corrected.

A number of the discussed reasons for poor data quality are in line with those mentioned in the literature, in section 3.5. Data problems related to sensor errors have gotten additional attention in this study, since the RFID system was the data source where most errors were found. These results are likely applicable to several companies that use RFID systems, as well as other sensor systems. Problems related to RFID data are also highlighted in the interview with MagnorVinduet, in section 2.2, where they claimed to have experienced several of the same issues with their previously implemented RFID technology. However, depending on the industry and the work at hand, the issues and their causes may differ. A summary of the identified issues, their root causes, data quality dimensions, impact on ML models, and suggested improvement actions are presented in Figure 34.

Moreover, data quality has a crucial impact on the performance of LTP. A number of issues were found to lower the data quality of the RFID data. The findings highlight inconsistent and missing shop floor scannings, missing values, and contradictory values. Sensor node errors, outdated data, lack of validation, poor system design, and human errors are some of the detected potential causes for poor data quality, which decreases the completeness, consistency, believability and accuracy of the data. This affects the ML models ability to detected patterns among the variables, due to erroneous relationship detection, missing records and faulty data. An RFID system provides available data, but not necessary reliable data. The fenestration industry has potential to successfully install and utilize an RFID system, but the results show that the data needs to be explored and analyzed in order to detect faults and improve the system. Implementation of strategies and procedures in the company to mitigate these challenges are essential in order to improve the data quality and ensure accurate and reliable LTP.

Detected problem	DQ dimension	Possible cause	ML affect	Improvement
Missing variation/values	Completeness	• Obsolete data • Missing data • Data covering too short time period	Incorrect prediction of new orders with other characteristics that are not captured by the training data.	Update training data on regular basis.
Inconsistent and missing shop floor scannings	Completeness, Consistency	• Multiple scans for a single component • Sensor nodes malfunctions • Poor sensor node quality • Calibration issues • RFID tag falls off • Operator's ignorance and negligence • Lack of validation • Network or communication errors • Components that exit and reenter production line	Incorrect data entries to ML model. Incorrect calculation of output variable, lead time, and system status variables, which are the most influential for precise lead time prediction.	Company procedures to verify the validity and consistency of the RFID data. Consistency threshold to benchmark the completeness of shop floor scanning. External help to check the correctness of the RFID system. Several RFID scanners.
Inconsistent shop floor scanning points after data cleaning	Completeness, Accuracy	• Incomplete data cleaning • Uniqueness constrains and functional dependency violation	Inconsistent values for comparable components that ought to have been equal reduces the ML models' capacity to detect patterns between the variables.	Improved data cleaning.
Contradictory values in dataset 1 and 2	Believeablity, Consistency	• Obsolete data in dataset 2 • Operator's ignorance and negligence in dataset 1 • Sensor nodes malfunctions in dataset 1 • Calibration issues in dataset 1 • Poor sensor node quality in dataset 1 • Network or communication errors in dataset 1	Without accurate and trustworthy data the ML models prediction of future lead times are likely to be inaccurate, since the model learn erroneous relationship among the features.	Company procedures for updating data records. Improved integration between data sources. See improvement for «Inconsistent and missing shop floor scannings».
Reduction of input dataset	Accuracy	• Data cleaning	Lower performance due to reduced training data.	Improve business and data understanding to possibly reduce the amount of removed data.

Figure 34: Summary of the detected data quality problems, dimensions, causes, affect on ML and improvement suggestions.

7.3 RQ3: How can CRISP-DM be applied to predict lead times in the MTO fenestration industry?

CRISP-DM has been identified in the literature as a popular process model to use when working with ML. CRISP-DM was established with the aim of developing an industry standard for how practitioners perform ML, and incorporates significant elements of the ML life-cycle (Clark 2018). The process model can be adapted to several problems and industries. As the authors know of, there has not been conducted a detailed description of the actions that must be taken for each step in the model within the context of LTP for the fenestration industry. To answer RQ 3 the authors have the authors have developed a process model that expands the CRISP-DM model and makes it applicable for ML-based LTP in the MTO fenestration industry. This was achieved by combining the findings from RQ1, RQ2, and by developing ML models with the RFID data from the case company, presented in section 6.3. The process model aids to ease the phase, for businesses with comparable traits, of implementing ML-based LTP. The steps that make up the CRIPS-DM model are discussed in this section, with the goal of identifying the required actions for each step in the MTO fenestration industry.

Business understanding

Gaining a proper business understanding is the first phase of CRISP-DM. It is important to get familiar with business objectives in order to convert them to ML problems (Meidan, Lerner, Hassoun et al. 2009). The business understanding gained in this thesis was achieved in two ways. Firstly, the most crucial source to business understanding was obtained through the case study, presented in section 4. This included an understanding of their market, production processes, and planning system. Additionally, detection of challenges faced by the company, as well as potential improvement areas, presented in section 4.5, was assesses in order to understand the firms objective. Qualitative methods were applied such as interviews, meetings, visit at their factory and questionnaires, several which were performed during the authors project thesis. Two additional companies in the fenestration industry, presented in section 2, were also interviewed to get a broader business understanding. In addition, literature was reviewed about the industry and LTP. The findings were presented in section 1.1 and 3.

The findings showed that the fenestration market is affected by increasing demand for customized products, increasing competition from low-cost countries and stricter government regulations for energy efficient windows. This pushes the companies to produce high-quality windows on-time and at lower cost. The degree of international competition is likely to vary depending on the investigated nation. The competition from low-cost countries, stated by the case company, is probably more applicable for fenestration companies in high labour cost countries, such as the interviewed companies from Norway. Government regulations for energy efficient windows can be both national and international, and therefore the degree of regulations will also vary between nations. On the other hand, this is an increasing focus all over the world, and all fenestration companies are likely to be affected in some degree.

In order to obtain high-quality windows at low cost the importance of production flexibility and manufacturing efficiency was mentioned as a crucial step. MTO fenestration companies are characterized by high level of product variety, which increases the planning and production complexity. As seen from the company interviews they have varying production layout, degree of digitization and planning systems. The lead time is affected by these factors. As could be seen, Gilje has the shortest lead time, 1-2 days, and linear layout, compared to the others where the lead time was approximately 1-2 weeks and functional layout. None of the companies could give a standard estimation of the lead time, since it varied for each product. A general characteristic for the companies was that the customer orders were placed several weeks or months in advance of delivery, which could increase the possibility of re-planning. All companies delivered a high variety of products which could be configured by the customer.

The high degree of customization in MTO fenestration companies result in several factors affecting the lead time. The detected lead time factors from the literature were presented in section 3.4, where system status features were highlighted to contain the most important information affecting the lead time, thereby WIP levels as the most influential, followed by bottlenecks and machine status. Some emphasize was also given to order characteristics, like product priority and processing times. From the questionnaire presented to Gilje, in section 4.5.1, it was highlighted that system status, specially bottlenecks, buffer load, product mix, changeover time, and utilization and downtime of the machines, affected the lead time at the factory. Gilje also noted that the type of product, number of components, number of glass panes, color and the number of colors had an effect on the lead time, since these factors initialize individual requirements for the order. The order characteristics Gilje focused on can be viewed as specific for the fenestration industry.

Similar to Gilje, both Nordic Door and MagnorVinduet mentioned bottleneck as one of the most influential lead time factors. Order characteristics were also highlighted as influential. As seen in the findings, a bigger emphasize to order characteristics was given by the fenestration companies than what was discovered in the literature. This may be because the system status has a large effect on lead time for all industries, but it depends on the product type how influential the order characteristics are. In addition to the previously identified factors, Nordic Door mentioned raw material delay as a potential lead time affecting factor, while MagnorVinduet mentioned setup times. Performing a proper business understanding is crucial in order to detect lead time affecting parameters, which can vary between companies. The factors can be identified by integrating findings from literature with findings from a case study.

Data understanding

In order to collect the pertinent data detected in the business understanding phase, relevant data sources must be identified. The necessary data for the task at hand was identified to be data about product characteristics and system status. In the case of Gilje, product characteristics was achieved through order data retrieved from their ERP system, and RFID data was utilized to retrieve information about the system status. As seen in the literature, RFID systems and other sensor systems have been identified as appropriate data sources, able to provide an extensive amount of production data. RFID data was used to extract system status variables; the buffer level, which in this case was the bottleneck and WIP level, amount of components in the system, and amount of components in a given series, as presented in section 5.1.2. The study showed that the RFID data could retrieve these features based on straightforward calculations, and as a result, it can be a useful data source for LTP. In the study of S. Huang, Guo et al. (2019), presented in section 3.4, they were able to calculate additional features from RFID data, such as the processing time at each machine since all machines had RFID scanner in inn-out pairs. This was not obtained in this study due to the inconsistent RFID scanning points of the few machines that had inn-out pairs. This shows that the retrievable system status features from RFID data is affected by the placement of the RFID scanners and is context dependent. In addition, machine related variables, such as utilization and downtime of the machines and changeover time, were also detected as useful input features for the ML model. This was not possible to obtain from the data at hand in this study. In a future implementation, machine data should be included if the data is obtainable.

A thorough descriptive analysis of the data is important in order to obtain an understanding of the data's nature and quality (Meidan, Lerner, Hassoun et al. 2009). Through an assessment of the data quality, the researchers were able to identify data issues and data limitations that could potentially affect the ML results, which was discussed in section 7.2. These detection's gave a threshold to decide what data to include and exclude in the subsequent steps of the CRISP-DM model. This is an important decision to make when working with ML, since appropriate data inputs are crucial in order to obtain accurate predictions. The necessity of a complete data quality assessment depends on the reliability and knowledge about the data source. In this thesis, the historical data had not been utilized by the company, and they had limited knowledge about the quality of the data. The need for a data quality assessment may increase when working with sensor data. Prior to the assessment, the authors found it challenging to identify the inconsistent RFID scanning points, but this might also be a result of the format in which the data was received. Therefore the necessity and the extent of a data quality assessment can be considered context specific.

Data preparation

The data preparation step involves the transformation of raw data into a cleansed dataset that can be used by a ML model. A detailed description of the applied data preparation was presented in section 5.1.2. Based on the findings and result from the study, six essential steps has been identified and performed; feature calculation, data cleaning, data encoding, feature extraction, data transformation and dimensionality reduction. The literature highlight the steps of data cleaning as removing or replacing missing values, duplicate data, invalid data and noise (Javatpoint 2023; Maharana, Mondal et al. 2022). As detected during the data quality assessment, RFID data may contain a lot of noise; incorrect scannings or system errors that lead to duplicates, outliers or other defects that affect the data quality. The target variable for the ML model was calculated from the RFID data, therefore removing noise and invalid data has received an extra emphasize in this study. Data cleaning was concentrated around removing data records with incomplete RFID scannings, duplicates, manually processed and reordered components. The need for these removals were detected during the business and data understanding phase. Outlier detection was also applied in order to decrease the impact of faulty data records that had not been identified in the prior cleaning. Feature extraction is context dependent, and a proper business and data understanding is needed in order to only remove columns that contain redundant information. Through communication with Gilje and the data quality assessment, features specific to Gilje, such as fastlane orders and lev_eno_uno tags, were removed during this phase.

By carrying out these data preparation steps, researchers may be able to resolve several data quality issues discovered during data understanding. Therefore, ensuring accurate and consistent data preparation is essential. The necessary data preparation steps are context specific, but several of the executed steps are widely applied when working with ML, such as dimensionality reduction with PCA and IQR outlier detection. In addition, this thesis has demonstrated the significance of removing missing values when working with RFID data. This is likely applicable to other sensor systems in several industries. As evidenced by the literature, a number of the observed flaws stem from issues with sensor data in general, not only RFID data. As discussed under "data understanding", it is possible that some of the data preparation steps would not have been necessary if a more trustworthy data source had been used, e.g. removing the large amount of data records with inconsistent scanning points. If the company uses the data source frequently for other planning tasks, a number of flaws may already be detected and improved.

Data modelling

The literature emphasizes the lack of a generalized model for LTP, making it necessary to examine various models to determine which one best suits the issue at hand (Kramer, C. Wagner et al. 2020). The literature review on ML for lead time prediction, section 3.4, was used to restrict the field of model candidates. MLP ANN and RF were identified as two common approaches used for LTP. In prior studies, both models have shown a superiority over other methods, such as LR. They were also proved to be applicable for the fenestration environment, with the ability to handle complex production processes, with independent routing and high order variety.

The first experiment, in section 6.3.1, was conducted in order to test the superiority of these models. The results were consistent with the literature, where the LR model was outperformed by the RF model and the MLP model without hyperparameter tuning. The superior performance of ANN and RF may be due to their ability to recognize non-linear relationships and manage intricate interactions between the input features. Since LR models only consider linear relationships, they may not be able to adequately represent the complexity found in MTO fenestration companies. Therefore, ANN and RF are considered to be more suited for this purpose.

Hyperparameter tuning has been mentioned in the literature as an important step during the modeling phase (Bardenet, Brendel et al. 2013). Based on the findings in the literature study, the authors decided that grid search and randomized search should be executed. Tuning of the parameters is crucial, since it helps optimize the performance of the ML model. The researcher may be able to fine-tune the model's behavior to improve accuracy and increase the overall efficiency of the ML model by adjusting the values of the hyperparameters. In section 6.3, different ML experiments were performed. In experiment 1, section 6.3.1, the MLP model and RF model were trained without hyperparameter tuning, while in experiment 2, section 6.3.2, hyperparameter tuning was performed on both of the models. The MLP model and the RF model performed better after hyperparameter tuning, according to the results. This was consistent with findings from the literature, which highlights the importance of hyperparameter tuning.

Evaluation

The evaluation of the models was done using the metrics specified in section 5.2.1. The metrics were selected based on the metrics that were most commonly mentioned in the literature for regression tasks.

The first experiment in section 6.3.1 also tested another theory from the literature; the importance of system status variables. Results from the experiment showed that an inclusion of the system status improved the model performance, which align with theory. Data that represent system status often correlate with the lead time and should therefore be included when performing LTP. This is also supported by the findings from the feature importance analysis in the same section. Findings demonstrate that the three system status features were precisely the features that had the greatest influence on lead time. This also shows the importance of including additional system status variables if it is obtainable, as discussed under "data understanding". A better performing model could be achieved if the shop floor load was further reflected, e.g. with machine data. The fourth most important feature was the component length. Although the length was significantly less important than the system status factors, this may represent the impact of the high amount of customization in MTO fenestration companies. In a more uniform production, the order characteristics are probably less influential. The number of "glass points", amount of frame and casement components had the same variable importance, which was expected since all depend on the number of glass panes in a window.

In experiment 2, see section 6.3.2, the MLP model and RF model were evaluated after hyperparameter tuning using the defined performance metrics. Based on the results displayed in the table, the RF model outperformed MLP across all performance metrics. The RF model's greater R^2 value indicates a superior ability to capture variation in the target variable. RF also showed improved predictive accuracy and lower errors, as evidenced by decreased MAE, MSE, and RMSE values. These benefits can be attributed to the robustness of the RF model to overfitting as well as its capability to handle complicated interactions and non-linear relationships in the data. MLP models, in contrast, rely on interconnected layers of neurons and call for precise hyperparameter tuning, making them more sensitive to the quantity and quality of the data. Therefore the low performance of the MLP model might be related to the poor data quality of the RFID data. There may still be inaccurate data records, as was mentioned in section 7.2, which has a direct impact on the target variable and the derived system status attributes, which have been shown to be the most influential factors. The RF model can be viewed as a more effective approach with overall superior model performance compared to the MLP for the given dataset.

Even though RF outperformed MLP in terms of results, the model still has improvement potential. The model's inability to account for all variation in the initial data was demonstrated by the distribution of predicted variables, which is presented in section 6.3.2. The poor data quality is a probable explanation for this as well, due to the required amount of reduction of the input dataset after data preparation. A dataset collected over a longer period of time should be taken into consideration for future exploration, which can reduce the issue of unnoticed variances in the dataset, and the market trend will be more accurately reflected. If the original data comprises several records, the dataset will also be less sensitive to extensive data cleaning.

Deployment

In the final phase of the ML-life cycle, the ML model may be integrated into a real-world production system. A deployment of the model can provide additional value to the company, and should align with the company needs, identified during the business understanding. However, as mentioned in section 2.4, the deployment focus in this thesis is on finding viable PPC application areas for LTP. A discussion of potential application areas was presented in section 7.1. The application areas most relevant for the fenestration industry were capacity planning, production monitoring and analysis. These results can be used by other companies operating in the same field as a basis for exploration of potential application areas that may be suitable for them. The detected application areas should be further investigated and verified through quantitative methods.

In summary, the CRISP-DM procedures has been applied to a case company from the MTO fenestration sector to investigate how LTP can be applied using ML within this industry. Business understanding was achieved through a case study of Gilje and a LR, with additional input from two other Norwegian fenestration companies. Through the business understanding, pertinent lead time affecting features were detected. In this thesis a data quality assessment was applied to achieve data understanding, and identify areas that need data preparation. Several data preparation techniques were applied, thereby calculation of system status features, which later proved to be the most influential features for the ML models. Two ML models were developed for LTP, RF and MLP ANN. In addition to proving the importance of system status variables, the developed models also outperformed the LR model used for comparison. After hyperparameter tuning RF proved to be the best model for LTP, in the investigated case company. The deployment potential for LTP in PPC was investigated in terms of the three interviewed MTO fenestration companies. The final process model is presented in Figure 35.

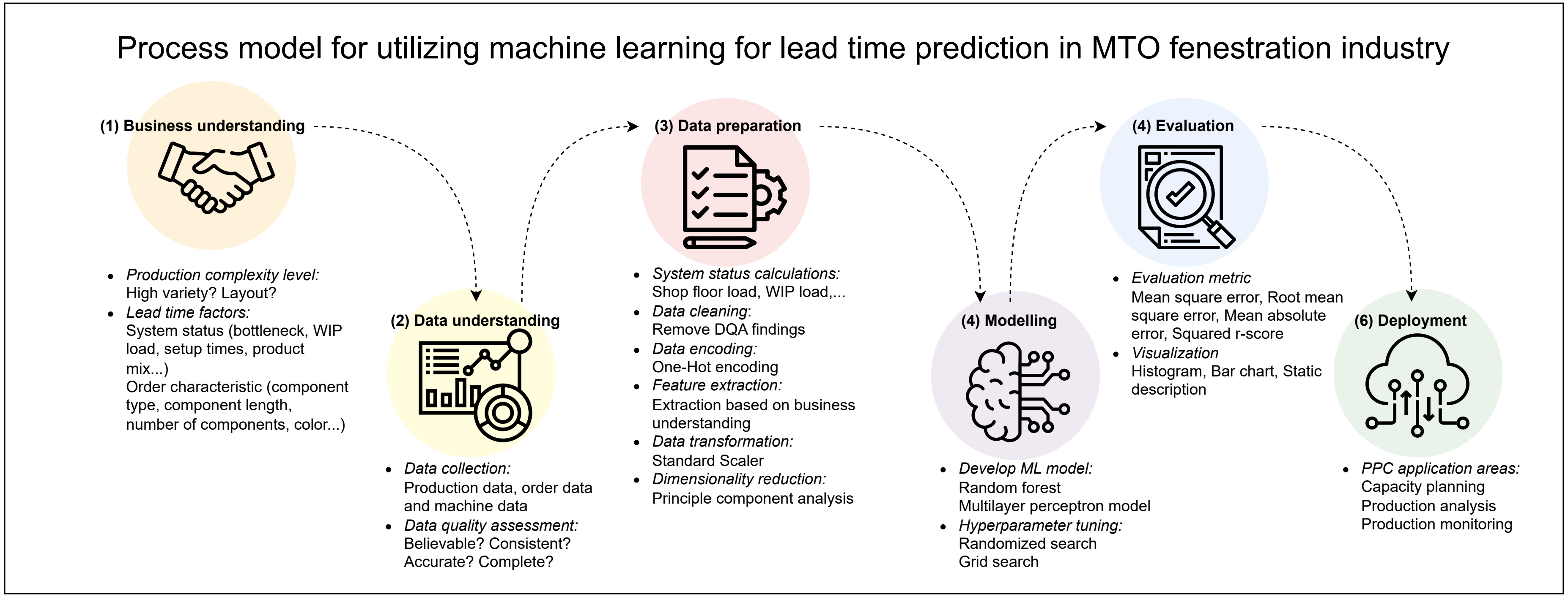


Figure 35: Process model for development of ML-based LTP in the MTO fenestration industry

8 Conclusions

In conclusion, the MTO fenestration industry is impacted by rising consumer demand for customized products and stricter national and international energy/climate restrictions, which increases the need for reliable and flexible PPC to meet the shifting market. Therefore, this thesis aims to contribute to increased understanding of how ML-based lead time prediction could be developed and integrated to support PPC in the MTO fenestration industry. RQ1, RQ2 and RQ3 were answered in section 7. A process model for LTP in the MTO fenestration industry was developed to ease the phase of implementing ML-based LTP. Special focus was given to the assessment of data quality and potential integration of LTP in the industry's planning approaches.

LTP can improve various areas of production planning and control in MTO manufactures. Through a literature review, capacity planning, production scheduling, due date setting, production analysis and production monitoring were detected as the most relevant application areas for LTP. Three MTO fenestration companies in Norway were interviewed to detect the transferability to the MTO fenestration industry, presented in section 6.1. The responses from the two larger companies were highly identical and listed the most pertinent LTP application areas as capacity planning, production monitoring, and production analysis. Today, the companies' capacity planning approach only takes into account one factor, which causes deviation between anticipated and actual production. By employing an approach based on LTP, the company could enhance the reliability of the capacity plan by better reflecting the real-world capacity. Predicting the remaining lead time for production monitoring could ease re-prioritization of orders during manufacturing and reduce delays due to insufficient planning. By utilizing LTP for production analysis, the companies saw potential for optimizing a number of production factors, including product mix and buffer level. The smaller company that was interviewed found LTP to offer minimal benefits across all application areas. As a smaller company, the necessity of digitizing PPC might be less than the cost of it, since it is easier to handle planning manually. Application of LTP for due date setting and production scheduling was not deemed important for any of the interviewed companies, because customers set the due dates and their scheduling methods are independent of lead times.

It is crucial to consider data quality when working with ML. Therefore a data quality assessment was conducted to identify data quality issues, root causes and improvement potential for the applied data for LTP. The data quality was defined according to four dimensions detected in the literature; completeness, consistency, believability and accuracy. The results, presented in section 6.2, reveled highly incomplete and inconsistent shop floor scannings', with only 60% of the components successfully navigating through all scanning points. Detected contradictory values between datasets reduced the believability of the data sources. The final input dataset for the ML model was reduced by 53% of the original data records. Potential causes for poor data quality was addressed from previous literature, and in communication with the case company. The main causes detected for the data quality problems were errors in the RFID system, missing values, obsolete data, and operators' ignorance and negligence. A discussion of potential improvement strategies for the company was also presented in section 7.2.

To evaluate models for LTP, two ML models were implemented. According to the literature a generalized model for LTP does not exist, hence several models needs to be compared. A literature review was conducted to detect prior ML models utilized for LTP. The two best performing models detected in the literature, the random forest and multilayer perceptron models, were further compared. A linear regression model was used as baseline performance in order to evaluate and compare the models. LR was outperformed by both RF and ANN. After hyperparameter tuning was applied, RF showed superior performance over MLP in all evaluation metrics.

The results from this thesis provided the foundation to develop a process model for predicting lead times using ML in the MTO fenestration industry. The well-known CRISP-DM framework was used as the base for the model, incorporating it's six steps. First, *business understanding* of the fenestration industry was achieved through the case company and additional interviews with other companies in the same sector. This provided an understanding of the necessary input features, system status and order characteristics. Further, *data understanding* was achieved from the data quality assessment and in communication with the case company. The data issues detected during the assessment was eliminated during *data preparation*, which provided a final input dataset for the ML models. Lastly, *modelling* and *evaluation* of RF and MLP was conducted with the input dataset developed in the prior step. As mentioned, RF provided the best results, and the feature importance reveled superiority of including system status features compared to order characteristics. The *deployment* was not put in practice in this thesis. However, potential applications areas were addressed and examined, and the obtained knowledge could be used for future deployment plans.

The thesis main contribution to theory is an extension of the existing CRISP-DM framework by developing an industry specific process model for ML-based LTP in the MTO fenestration industry. The theoretical contribution provides a structured approach that integrates domain-specific knowledge, data quality assessment, quantitative modelling, and PPC deployment areas into the ML implementation process for LTP. The practical contribution is a guide for implementing ML-based LTP. This can contribute to a shared understanding between data scientists and production managers, which can assist companies in effectively integrating LTP into their PPC processes, thereby enhancing operational planning and decision-making. As a step towards enhancing fenestration companies' planning and decision-making processes, this can facilitate the companies to better handle future energy regulations, and overall reduce their environmental footprint by optimizing the production.

8.1 Limitations and Future work

The developed process model in this thesis was mainly based on the findings from one MTO fenestration company in Norway, Gilje, with some additional input from two other companies in the same sector. The three fenestration companies vary both in size, production process and digitization level, which may indicate a great variability within the fenestration industry. This shows that Gilje only represent one part of the fenestration industry. It is likely in the upper echelon when it comes to digitization as the company has integrated an RFID system in the production, as well as several lean practices. Because of this, they may have a more standardized production than other companies in the same sector. Companies with more traditional production processes can have more complex routings, and other factors influencing the lead time. Since all data-related findings were dependent on the data from Gilje, other companies may find different input features and ML models to be more suitable. This limits the generalizability of the process model. Future research should consider replicating this study with other data sources and input features within different fenestration environments to establish the robustness and generalizability of the process models. It may be pertinent to test and assess the process model on companies that use more conventional production methods in order to represent a wider range of the industry.

Additionally, future researchers could incorporate several system status variables that were not included in this study, thereby machine data, to verify the discussed impact. In order to collect this data, other data sources could be of interest, such as various sensor systems or PPC data management systems. By employing other data sources, additional data quality issues could be identified. As mentioned, it could be valuable to explore several ML models for the task of LTP. The choice of ML algorithms used in this study was based on their established success in previous research. However, the field of ML is rapidly evolving, and new algorithms and techniques may emerge that could potentially improve LTP even further.

As discussed in section 7.1, the results for PPC application areas were dispersed between the interviewed companies. Several companies in the MTO fenestration industry, in different countries, should be considered in future research to provide clearer guidelines for the applicable PPC areas. In addition, quantitative methods should be applied to investigate real-world integration of LTP in the detected areas. This can further verify the applicability of adapting LTP in MTO fenestration industries PPC system.

References

Ahmarofi, Ahmad Afif, Razamin Ramli and Norhaslinda Zainal Abidin (2017). 'Predicting completion time for production line in a supply chain system through artificial neural networks'. In: *International Journal of Supply Chain Management* 6.3, pp. 82–90.

Alenezi, Abdulrahman, Scott A Moses and Theodore B Trafalis (2008). 'Real-time prediction of order flowtimes using support vector regression'. In: *Computers & Operations Research* 35.11, pp. 3489–3503.

Alwan, Ahmed Abdulhasan et al. (2022). 'Data quality challenges in large-scale cyber-physical systems: A systematic review'. In: *Information Systems* 105, p. 101951.

Asadzadeh, SM, Ali Azadeh and A Ziaeifar (2011). 'A neuro-fuzzy-regression algorithm for improved prediction of manufacturing lead time with machine breakdowns'. In: *Concurrent Engineering* 19.4, pp. 269–281.

Atik, Ceren, Recen Alp Kut and Safak Birol (2021). 'Estimating Lead Time Using Machine Learning Algorithms: A Case Study by a Textile Company'. In: *2021 Innovations in Intelligent Systems and Applications Conference (ASYU)*. IEEE, pp. 1–5.

Backus, Phillip et al. (2006). 'Factory cycle-time prediction with a data-mining approach'. In: *IEEE Transactions on Semiconductor Manufacturing* 19.2, pp. 252–258.

Bardenet, Rémi et al. (2013). 'Collaborative hyperparameter tuning'. In: *International conference on machine learning*. PMLR, pp. 199–207.

Bender, Janek and Jivka Ovtcharova (2021). 'Prototyping machine-learning-supported lead time prediction using AutoML'. In: *Procedia Computer Science* 180, pp. 649–655.

Bender, Janek, Martin Trat and Jivka Ovtcharova (2022). 'Benchmarking automl-supported lead time prediction'. In: *Procedia Computer Science* 200, pp. 482–494.

Benesty, Jacob, Jingdong Chen and Yiteng Huang (2008). 'On the importance of the Pearson correlation coefficient in noise reduction'. In: *IEEE Transactions on Audio, Speech, and Language Processing* 16.4, pp. 757–765.

Benjaoran, Vacharapoom and Nashwan Dawood (2005). 'A case study of artificial intelligence planner for make–to–order precast concrete production planning'. In: *Computing in Civil Engineering (2005)*, pp. 1–10.

Bisong, Ekaba and Ekaba Bisong (2019). 'NumPy'. In: *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*, pp. 91–113.

Blackston, Michelle (2022). *why fenestration is a crucial part of a sustainable future*. Accessed on 05 21, 2023. URL: https://www.windowanddoor.com/blog/why-fenestration-crucial-part-sustainable-future.

Buer, Sven-Vegard et al. (2018). 'Strategic fit of planning environments: towards an integrated framework'. In: *Information Systems, Logistics, and Supply Chain: 6th International Conference, ILS 2016, Bordeaux, France, June 1–4, 2016, Revised Selected Papers 6*. Springer, pp. 77–92.

Burggräf, Peter et al. (2020). 'Approaches for the prediction of lead times in an engineer to order environment—A systematic review'. In: *IEEE Access* 8, pp. 142434–142445.

Cañas, Héctor et al. (2022). 'A conceptual framework for smart production planning and control in Industry 4.0'. In: *Computers & Industrial Engineering* 173, p. 108659.

Cao, Huaqing and Xiaofen Ji (2021). 'Prediction of Garment Production Cycle Time Based on a Neural Network'. In: *Fibres & Textiles in Eastern Europe* 1 (145, pp. 8–12.

Chai, Tianfeng and Roland R Draxler (2014). 'Root mean square error (RMSE) or mean absolute error (MAE)'. In: *Geoscientific model development discussions* 7.1, pp. 1525–1534.

Chapman, Stephen N (2006). *The fundamentals of production planning and control.* Pearson/Prentice Hall Upper Saddle River, NJ, USA.

Chen, Mengjie et al. (2012). 'Survey on data quality'. In: *2012 World Congress on Information and Communication Technologies*. IEEE, pp. 1009–1013.

Chen, T (2009). 'A fuzzy-neural knowledge-based system for job completion time prediction and internal due date assignment in a wafer fabrication plant'. In: *International Journal of Systems Science* 40.8, pp. 889–902.

Chicco, Davide, Matthijs J Warrens and Giuseppe Jurman (2021). 'The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation'. In: *PeerJ Computer Science* 7, e623.

Chien, Chen-Fu, Chia-Yu Hsu and Chih-Wei Hsiao (2012). 'Manufacturing intelligence to forecast and reduce semiconductor cycle time'. In: *Journal of Intelligent Manufacturing* 23, pp. 2281–2294.

Choueiri, Alexandre Checoli et al. (2020). 'An extended model for remaining time prediction in manufacturing systems using process mining'. In: *Journal of Manufacturing Systems* 56, pp. 188–201.

Cichy, Corinna and Stefan Rass (2019). 'An overview of data quality frameworks'. In: *IEEE Access* 7, pp. 24634–24648.

Clark, Andrew (2018). 'The machine learning audit-CRISP-DM Framework'. In: *Isaca Journal* 1, pp. 42–47.

Corrales, David Camilo, Juan Carlos Corrales and Agapito Ledezma (2018). 'How to address the data quality issues in regression models: a guided process for data cleaning'. In: *Symmetry* 10.4, p. 99.

Corrales, David Camilo, Agapito Ledezma and Juan Carlos Corrales (2018). 'From theory to practice: A data quality framework for classification tasks'. In: *Symmetry* 10.7, p. 248.

Cos Juez, Francisco Javier de et al. (2010). 'Analysis of lead times of metallic components in the aerospace industry through a supported vector machine model'. In: *Mathematical and computer modelling* 52.7-8, pp. 1177–1184.

Council, World Green Building (2019). 'Bringing Embodied Carbon Upfront: Coordinated action for the building and construction sector to tackle embodied carbon'. In.

Cox, James F and John H Blackstone (2002). *APICS dictionary.* APICS Educational Society for Resource Management.

Dalinina, Ruslana (2017). *Introduction to Correlation. Oracle+ DataScience. com.*

Eckerson, Wayne W (2002). 'Data quality and the bottom line'. In: *TDWI Report, The Data Warehouse Institute*, pp. 1–32.

El Naqa, Issam and Martin J Murphy (2015). *What is machine learning?* Springer.

Eppler, Martin J (2006). *Managing information quality: Increasing the value of information in knowledge-intensive products and processes.* Springer Science & Business Media.

EU (2023). *Energy performance of buildings directive.* Accessed on 05 21, 2023. URL: https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive_en#documents.

Fang, Weiguang et al. (2020). 'Big data driven jobs remaining time prediction in discrete manufacturing system: a deep learning-based approach'. In: *International Journal of Production Research* 58.9, pp. 2751–2766.

Fussenegger, Paul and Niklas Lange (2022). *Explainable Machine Learning for Lead Time Prediction: A Case Study on Explainability Methods and Benefits in the Pharmaceutical Industry.*

Gacek, Stanisław (2018). 'Due date assignment using neural networks for standard products in small batch and multi assortment make-to-order company'. In: *Proceedings of the Carpathian Logistics Congress, Prague, Czech Republic*, pp. 3–5.

Géron, Aurélien (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow.* " O'Reilly Media, Inc."

Gerring, John (2004). 'What is a case study and what is it good for?' In: *American political science review* 98.2, pp. 341–354.

— (2006). *Case study research: Principles and practices.* Cambridge university press.

Glowalla, Paul et al. (2014). 'Process-driven data quality management–An application of the combined conceptual life cycle model'. In: *2014 47th Hawaii International Conference on System Sciences.* IEEE, pp. 4700–4709.

Grolinger, Katarina et al. (2014). 'Challenges for mapreduce in big data'. In: *2014 IEEE world congress on services.* IEEE, pp. 182–189.

Gudivada, Venkat, Amy Apon and Junhua Ding (2017). 'Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations'. In: *International Journal on Advances in Software* 10.1, pp. 1–20.

Gyulai, Dávid, András Pfeiffer, Júlia Bergmann et al. (2018). 'Online lead time prediction supporting situation-aware production control'. In: *Procedia CIRP* 78, pp. 190–195.

Gyulai, Dávid, András Pfeiffer, Gábor Nick et al. (2018). 'Lead time prediction in a flow-shop environment with analytical and machine learning approaches'. In: *IFAC-PapersOnLine* 51.11, pp. 1029–1034.

Hasan, Hasmarina and Nooritawati Md Tahir (2010). 'Feature selection of breast cancer based on principal component analysis'. In: *2010 6th International Colloquium on Signal Processing & its Applications*. IEEE, pp. 1–4.

Haynes, Paula J, Marilyn M Helms and Robert S Boothe (1991). 'Rethinking the manufacturing focus: An overlooked strategic'. In: *SAM Advanced Management Journal* 56.4, p. 34.

Hendry, LC and BG Kingsman (1993). 'Customer enquiry management: part of a hierarchical system to control lead times in make-to-order companies'. In: *Journal of the operational research society* 44, pp. 61–70.

Hendry, Linda C and BG Kingsman (1989). 'Production planning systems and their applicability to make-to-order companies'. In: *European journal of operational research* 40.1, pp. 1–15.

Hiller, T, L Deipenwisch and P Nyhuis (2022). 'Systemising Data-driven Methods for Predicting Throughput Time within Production Planning & Control'. In: *2022 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*. IEEE, pp. 0716–0721.

Hsu, SY and DY Sha (2004). 'Due date assignment using artificial neural networks under different shop floor control strategies'. In: *International Journal of Production Research* 42.9, pp. 1727–1745.

Huang, Shaohua et al. (2019). 'A two-stage transfer learning-based deep learning approach for production progress prediction in IoT-enabled manufacturing'. In: *IEEE Internet of Things Journal* 6.6, pp. 10627–10638.

Hung, Yi-Feng, Chuan-Che Huang and Ying Yeh (2013). 'Real-time capacity requirement planning for make-to-order manufacturing with variable time-window orders'. In: *Computers & Industrial Engineering* 64.2, pp. 641–652.

Hvolby, Hans-Henrik and Anders Thorstenson (2001). 'Indicators for performance measurement in small and medium-sized enterprises'. In: *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* 215.8, pp. 1143–1146.

Iansiti, Marco and Karim R Lakhani (2014). 'Digital ubiquity:: How connections, sensors, and data are revolutionizing business'. In: *Harvard business review* 92.11, p. 19.

Jain, Abhinav et al. (2020). 'Overview and importance of data quality for machine learning tasks'. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 3561–3562.

Javatpoint (2023). *Machine learning life cycle*. URL: https://www.javatpoint.com/machine-learning-life-cycle (visited on 20th Mar. 2023).

Jayaweera, CD and N Aziz (2018). 'Reliability of principal component analysis and Pearson correlation coefficient, for application in artificial neural network model development, for water treatment plants'. In: *IOP Conference Series: Materials Science and Engineering*. Vol. 458. 1. IOP Publishing, p. 012076.

Jonsson, Patrik and Stig-Arne Mattsson (2003). ‘The implications of fit between planning environments and manufacturing planning and control methods’. In: *International Journal of Operations & Production Management.*

Juddoo, Suraj (2015). ‘Overview of data quality challenges in the context of Big Data’. In: *2015 International Conference on Computing, Communication and Security (ICCCS).* IEEE, pp. 1–9.

Kader, Shahidul and Maeen Md Khairul Akter (2014). ‘Analysis of the factors affecting the lead time for export of readymade apparels from Bangladesh; proposals for strategic reduction of lead time’. In: *European Scientific Journal* 10.33.

Kapulin, DV and PA Russkikh (2020). ‘Analysis and improvement of production planning within small-batch make-to-order production’. In: *Journal of Physics: Conference Series.* Vol. 1515. 2. IOP Publishing, p. 022072.

Kramer, Kathrin Julia, Carsten Wagner and Matthias Schmidt (2020). ‘Machine learning-supported planning of lead times in job shop manufacturing’. In: *Advances in Production Management Systems. The Path to Digital Transformation and Innovation of Production Management Systems: IFIP WG 5.7 International Conference, APMS 2020, Novi Sad, Serbia, August 30–September 3, 2020, Proceedings, Part I.* Springer, pp. 363–370.

Krogh, Anders (2008). ‘What are artificial neural networks?’ In: *Nature biotechnology* 26.2, pp. 195–197.

Kusiak, Andrew (2017). ‘Smart manufacturing must embrace big data’. In: *Nature* 544.7648, pp. 23–25.

Laranjeiro, Nuno, Seyma Nur Soydemir and Jorge Bernardino (2015). ‘A survey on data quality: classifying poor data’. In: *2015 IEEE 21st Pacific rim international symposium on dependable computing (PRDC).* IEEE, pp. 179–188.

Lee, Gyu M and Xuehong Gao (2021). ‘A hybrid approach combining fuzzy C-means-based genetic algorithm and machine learning for predicting job cycle times for semiconductor manufacturing’. In: *Applied Sciences* 11.16, p. 7428.

Li, Minqi et al. (2015). ‘Simulation-based experimental design and statistical modeling for lead time quotation’. In: *Journal of Manufacturing Systems* 37, pp. 362–374.

Liang, Feng (2010). ‘Cycle time estimation approach based on mathematical programming and statistical regression’. In: *2010 IEEE 17Th International Conference on Industrial Engineering and Engineering Management.* IEEE, pp. 742–745.

Lim, Zhong Heng, Umi Kalsom Yusof and Haziqah Shamsudin (2019). ‘Manufacturing lead time classification using support vector machine’. In: *Advances in Visual Informatics: 6th International Visual Informatics Conference, IVIC 2019, Bangi, Malaysia, November 19–21, 2019, Proceedings 6.* Springer, pp. 268–278.

Lindström, Veronica et al. (2023). ‘Data quality issues in production planning and control–Linkages to smart PPC’. In: *Computers in Industry* 147, p. 103871.

Lingitz, Lukas et al. (2018). 'Lead time prediction using machine learning algorithms: A case study by a semiconductor manufacturer'. In: *Procedia Cirp* 72, pp. 1051–1056.

Lu, S.C.H., D. Ramaswamy and P.R. Kumar (1994). 'Efficient scheduling policies to reduce mean and variance of cycle-time in semiconductor manufacturing plants'. In: *IEEE Transactions on Semiconductor Manufacturing* 7.3, pp. 374–388. DOI: 10.1109/66.311341.

Maćkiewicz, Andrzej and Waldemar Ratajczak (1993). 'Principal components analysis (PCA)'. In: *Computers & Geosciences* 19.3, pp. 303–342.

Maharana, Kiran, Surajit Mondal and Bhushankumar Nemade (2022). 'A review: Data pre-processing and data augmentation techniques'. In: *Global Transitions Proceedings*.

Mansouri, Taha et al. (Nov. 2021). 'IoT Data Quality Issues and Potential Solutions: A Literature Review'. In: *The Computer Journal* 66.3, pp. 615–625.

Meidan, Yair, Boaz Lerner, Michael Hassoun et al. (2009). 'Data mining for cycle time key factor identification and prediction in semiconductor manufacturing'. In: *IFAC Proceedings Volumes* 42.4, pp. 217–222.

Meidan, Yair, Boaz Lerner, Gad Rabinowitz et al. (2011). 'Cycle-time key factor identification and prediction in semiconductor manufacturing using machine learning and data mining'. In: *IEEE transactions on semiconductor manufacturing* 24.2, pp. 237–248.

Mestry, Siddharth, Purushothaman Damodaran and Chin-Sheng Chen (2011). 'A branch and price solution approach for order acceptance and capacity planning in make-to-order operations'. In: *European Journal of Operational Research* 211.3, pp. 480–495.

Mezzogori, Davide, Giovanni Romagnoli and Francesco Zammori (2019). 'Deep learning and WLC: how to set realistic delivery dates in high variety manufacturing systems'. In: *IFAC-PapersOnLine* 52.13, pp. 2092–2097.

— (2021). 'Defining accurate delivery dates in make to order job-shops managed by workload control'. In: *Flexible Services and Manufacturing Journal* 33.4, pp. 956–991.

Mikati, Nabil (2010). 'Dependence of lead time on batch size studied by a system dynamics model'. In: *International Journal of Production Research* 48.18, pp. 5523–5532.

Misiopecki, Cezary et al. (2018). 'Thermal modeling and investigation of the most energy-efficient window position'. In: *Energy and Buildings* 158, pp. 1079–1086.

Mori, Junichi and Vladimir Mahalec (2015). 'Planning and scheduling of steel plates production. Part I: Estimation of production times via hybrid Bayesian networks for large domain of discrete variables'. In: *Computers & Chemical Engineering* 79, pp. 113–134.

Mourtzis, Dimitris et al. (2014). 'Knowledge-based estimation of manufacturing lead time for complex engineered-to-order products'. In: *Procedia CIRP* 17, pp. 499–504.

Muñoz, Edrisi et al. (2012). 'Operational, Tactical and strategical integration for enterprise decision-making'. In: *Computer Aided Chemical Engineering*. Vol. 30. Elsevier, pp. 397–401.

Muralidharan, K (2010). 'A note on transformation, standardization and normalization'. In: *IUP J. Oper. Manag* 9, pp. 116–122.

Murphy, Rory et al. (2019). ‘Machine learning technologies for order flowtime estimation in manufacturing systems’. In: *Procedia CIRP* 81, pp. 701–706.

Oldebråten, Sylvia (2017). ‘Information Utilisation in the Planning Process of Suppliers in High-Variety, Low-Volume Supply Chains’. MA thesis. NTNU.

Oluyisola, Olumide Emmanuel (2021). ‘Towards smart production planning and control: Frameworks and case studies investigating the enhancement of production planning and control using internet-of-things, data analytics and machine learning’. In.

Öztürk, Atakan, Sinan Kayalıgil and Nur E Özdemirel (2006). ‘Manufacturing lead time estimation using data mining’. In: *European Journal of Operational Research* 173.2, pp. 683–700.

Parente, Manuel et al. (2020). ‘Production scheduling in the context of Industry 4.0: review and trends’. In: *International Journal of Production Research* 58.17, pp. 5401–5431.

Pathak, Vibha, Bijayini Jena and Sanjay Kalra (2013). ‘Qualitative research’. In: *Perspectives in clinical research* 4.3.

Pedregosa, Fabian et al. (2011). ‘Scikit-learn: Machine learning in Python’. In: *the Journal of machine Learning research* 12, pp. 2825–2830.

Perez, Alfonso Alexander (2014). ‘The value of accurate automated data collection to manufacturing’. PhD thesis. Massachusetts Institute of Technology.

Pfeiffer, András et al. (2016). ‘Manufacturing lead time estimation with the combination of simulation and statistical learning methods’. In: *Procedia Cirp* 41, pp. 75–80.

Pham, Duc T and Ashraf A Afify (2005). ‘Machine-learning techniques and their applications in manufacturing’. In: *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* 219.5, pp. 395–412.

Pipino, Leo L, Yang W Lee and Richard Y Wang (2002). ‘Data quality assessment’. In: *Communications of the ACM* 45.4, pp. 211–218.

Rajasekar, S, P Philominathan and V Chinnathambi (2006). ‘Research Methodology’. eng. In.

Regjeringen (2001). *Business and industry in Norway - The mechanical-engineering industry.* Accessed on 05 21, 2023. URL: https://www.regjeringen.no/en/dokumenter/Business-and-industry-in-Norway---The-mechanical-engineering-industry/id419355/.

Ren, Samuel Ching Xin et al. (2022). ‘Intelligent Manufacturing Planning System Using Dispatch Rules: A Case Study in Roofing Manufacturing Industry’. In: *Applied Sciences* 12.13, p. 6499.

Rodriguez-Galiano, V et al. (2015). ‘Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines’. In: *Ore Geology Reviews* 71, pp. 804–818.

Russell, Stuart J (2010). *Artificial intelligence a modern approach.* Pearson Education, Inc.

Sabbaghi, Asghar and Ganesh Vaidyanathan (2008). ‘Effectiveness and efficiency of RFID technology in supply chain management: strategic values and challenges’. In: *Journal of theoretical and applied electronic commerce research* 3.2, pp. 71–81.

Sagegg, Odd Jøran (2020). *ERP systems for manufacturing supply chains : applications, configuration, and performance.* eng. Boca Raton.

Samie, Farzad, Lars Bauer and Jörg Henkel (2019). 'From cloud down to things: An overview of machine learning in internet of things'. In: *IEEE Internet of Things Journal* 6.3, pp. 4921–4934.

Schkarin, Tatjana and Alexander Dobhan (2022). 'Prerequisites for Applying Artificial Intelligence for Scheduling in Small-and Medium-sized Enterprises.' In: *ICEIS (1)*, pp. 529–536.

Schneckenreither, Manuel, Stefan Haeussler and Christoph Gerhold (2021). 'Order release planning with predictive lead times: a machine learning approach'. In: *International Journal of Production Research* 59.11, pp. 3285–3303.

Schröer, Christoph, Felix Kruse and Jorge Marx Gómez (2021). 'A systematic literature review on applying CRISP-DM process model'. In: *Procedia Computer Science* 181, pp. 526–534.

Schuh, Günther, Andreas Gützlaff, Frederick Sauermann, Oliver Kaul et al. (2020). 'Databased prediction and planning of order-specific transition times'. In: *Procedia CIRP* 93, pp. 885–890.

Schuh, Günther, Andreas Gützlaff, Frederick Sauermann and Theresa Theunissen (2020). 'Application of time series data mining for the prediction of transition times in production'. In: *Procedia CIRP* 93, pp. 897–902.

Shah, Muhammad Izhar, Muhammad Faisal Javed and Taher Abunama (2021). 'Proposed formulation of surface water quality and modelling using gene expression, machine learning, and regression techniques'. In: *Environmental Science and Pollution Research* 28, pp. 13202–13220.

Sibanda, Nokuthaba (2009). 'Quantitative research'. In: *Wellington: Victoria University*.

Sjödin, David R et al. (2018). 'Smart Factory Implementation and Process Innovation: A Preliminary Maturity Model for Leveraging Digitalization in Manufacturing Moving to smart factories presents specific challenges that can be addressed through a structured approach focused on people, processes, and technologies.' In: *Research-technology management* 61.5, pp. 22–31.

SSB (2023). *Statistisk Sentralbyrå: Produksjonsindeks for bygge- og anleggsvirksomhet*. Accessed on 05 21, 2023. URL: https://www.ssb.no/statbank/table/13431/tableViewLayout1/.

Stevenson, Mark, Linda C Hendry and Brian G Kingsman (2005). 'A review of production planning and control: the applicability of key concepts to the make-to-order industry'. In: *International journal of production research* 43.5, pp. 869–898.

Strong, Diane M, Yang W Lee and Richard Y Wang (1997). 'Data quality in context'. In: *Communications of the ACM* 40.5, pp. 103–110.

Susanto, Steven, Prianggada Indra Tanaya and Adhi Sudadi Soembagijo (2012). 'Formulating standard product lead time at a textile factory using artificial neural networks'. In: *2012 2nd International Conference on Uncertainty Reasoning and Knowledge Engineering*. IEEE, pp. 99–104.

Szaller, Ádám et al. (2018). 'Real-time prediction of manufacturing lead times in complex production environments'. In.

Taleb, Ikbal et al. (2016). 'Big data quality: A quality dimensions evaluation'. In: *2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)*. IEEE, pp. 759–765.

Tenhiälä, Antti (2011). 'Contingency theory of capacity planning: The link between process types and planning methods'. In: *Journal of Operations Management* 29.1-2, pp. 65–77.

Thürer, Matthias et al. (2014). 'Lean control for make-to-order companies: Integrating customer enquiry management and order release'. In: *Production and Operations Management* 23.3, pp. 463–476.

Valarmathi, R and T Sheela (2021). 'Heart disease prediction using hyper parameter optimization (HPO) tuning'. In: *Biomedical Signal Processing and Control* 70, p. 103033.

Vinutha, HP, B Poornima and BM Sagar (2018). 'Detection of outliers using interquartile range technique from intrusion dataset'. In: *Information and Decision Sciences: Proceedings of the 6th International Conference on FICTA*. Springer, pp. 511–518.

Vollmann, T et al. (2004). *Manufacturing planning and control systems for supply chain management: the definitive guide for professionals*. Vol. 5. McGraw-Hill Professional.

Wang, Chuang and Pingyu Jiang (2019). 'Deep neural networks based order completion time prediction by using real-time job shop RFID data'. In: *Journal of Intelligent Manufacturing* 30, pp. 1303–1318.

Wang, Junliang and Jie Zhang (2016). 'Big data analytics for forecasting cycle time in semiconductor wafer fabrication system'. In: *International Journal of Production Research* 54.23, pp. 7231–7244.

Wang, Junliang, Jie Zhang and Xiaoxi Wang (2018). 'A data driven cycle time prediction with feature selection in a semiconductor wafer fabrication system'. In: *IEEE Transactions on Semiconductor Manufacturing* 31.1, pp. 173–182.

Wuest, Thorsten et al. (2016). 'Machine learning in manufacturing: advantages, challenges, and applications'. In: *Production & Manufacturing Research* 4.1, pp. 23–45.

Yamashiro, Hirochika and Hirofumi Nonaka (2021). 'Estimation of processing time using machine learning and real factory data for optimization of parallel machine scheduling problem'. In: *Operations Research Perspectives* 8, p. 100196.

Yin, Robert K (2009). *Case study research: Design and methods*. Vol. 5. sage.

Zhang, Rui and Cheng Wu (2012). 'A hybrid local search algorithm for scheduling real-world job shops with batch-wise pending due dates'. In: *Engineering Applications of Artificial Intelligence* 25.2, pp. 209–221.

Zijm, Willem HM and R Buitenhek (1996). 'Capacity planning and lead time management'. In: *International Journal of Production Economics* 46, pp. 165–179.

NTNU

Norwegian University of
Science and Technology