O.G. van Spengler

Two phase implementation of a machine learning (ML) model in demand forecasting for organisations

TU Delft

TWO PHASE IMPLEMENTATION OF A MACHINE LEARNING (ML) MODEL IN DEMAND FORECASTING FOR ORGANISATIONS

A thesis submitted to the Delft University of Technology in partial fulfillment of the requirements for the degree of

Master of Science
in Complex Systems Engineering and Management
Track: Transport and Logistics

by

O.G. (Olivier) van Spengler
Student ID: 5511011

February 2023

O.G. (Olivier) van Spengler: *Two phase implementation of a machine learning (ML) model in demand forecasting for organisations* (2023)

Faculty of Technology, Policy and Management
Delft University of Technology

Thesis committee: Prof. dr. ir. L.A. (Lóránt) Tavasszy, TU Delft, chair
Dr. J.H.R. (Ron) van Duin, TU Delft, first supervisor
Dr. Y. (Aaron) Ding, TU Delft, second supervisor
H.J. (Henk Jan) Jonge Poerink, Besi Apac., external supervisor
A. (Andreas) Schopper, Besi Switzerland AG, external supervisor

PREFACE

This thesis was written for BE Semiconductor Industries and the Delft University of Technology for the completion of the master Complex Systems Engineering & Management with the specialisation Transport & Logistics.

The research stems from interviews conducted with employees at all levels of the company Besi Apac. in Kuala Lumpur. Besi is a multinational company with Dutch origins that designs and manufactures semiconductor equipment. From the interviews a theme emerged that was the constantly altering sales forecast to which all production and procurement actions were tied. This issue could be tackled in various ways; however, the right way to fix this was the overlap of maximum result while operating within the confines of what a master thesis at the Delft University of Technology demands.

The study Complex Systems Engineering & Management is part of the Technology, Policy & Management faculty at the Delft University of Technology. The focus of the study is to learn operate in complex socio-technical environments to implement technical solutions. The approach is to look at all the facets surrounding the technology such as regulations, stakeholders, processes, etc. These aspects returned in this thesis to form the graduation project for this master.

EXECUTIVE SUMMARY

Demand forecasting plays a critical role in organizational planning, encompassing inventory management, capacity allocation, and financial decision-making. However, achieving accurate forecasts can be challenging, particularly in industries characterized by high demand volatility, such as semiconductor assembly equipment manufacturing, exemplified by Besi. Leveraging machine learning (ML) techniques presents a promising solution for effectively forecasting the seasonal and cyclical demand fluctuations experienced by Besi.

Besi, full name BE Semiconductor Industries N.V., is an multinational semiconductor equipment manufacturer which originates from The Netherlands. The company was founded in 1995 by Richard Blickman and now has operations in, among other countries, China, Switzerland and Malaysia. Besi develops leading edge assembly processes and equipment for leadframe, substrate and wafer level packaging applications in a wide range of end-user markets including electronics, mobile internet, cloud server, computing, automotive, industrial, LED and solar energy.

This paper investigates the challenges organizations face when implementing ML models into their demand forecasting processes, aiming to design a framework and implementation plan to guide organizations in adopting ML techniques. The research methodology employed is design science research (DSR), which focuses on developing and validating new designs within existing systems. The paper follows the iterative steps of DSR, including problem identification and motivation, objective definition, design and development, demonstration, evaluation, and communication. This iterative process facilitates collaboration with literature and industry experts to design a practical solution.

The study draws on literature research and exploratory discussions with Besi employees, emphasizing five key areas for investigation: the current forecasting process, existing forecasting techniques, organizational requirements, limitations, and input-output considerations. The findings highlight that Besi, similar to other organizations, employs a multi-layered forecasting process, with the most effective layer for implementing improvements and ML models being the initial forecast. Additionally, Besi predominantly relies on judgment-based forecasting techniques, making the implementation of a neutral ML tool necessary to create a hybrid forecasting system that mitigates human bias. Besi possesses the necessary prerequisites for effective ML techniques, such as clean and abundant data, but lacks the expertise required to construct and implement accurate models. Furthermore, Besi desires a neutral model that counters human bias and inputs historical monthly sales data, with the output expressed as total monthly sales.

To facilitate successful implementation within Besi's forecasting chain, several aspects are explored: the process framework (including integration, monitoring, updating, forecasting, and communication), peripheral considerations (e.g., legal and end-user trust), and the dashboard. The process framework is designed based on the existing forecasting process at Besi, incorporating the ML model, a dashboard, and revised information flows. The steps align with literature recommendations for ML-based forecasting, indicating that initial implementation is expected to face minimal resistance, monitoring is an ongoing task for the forecaster, updating involves improving the model and frequent training with new data, forecasting remains with the same personnel but incorporates additional information sources, and communication remains unchanged.

Peripheral matters, such as regulations and end-user trust, are limited in their impact, with research indicating that no laws impede ML model implementation, while a dashboard can enhance trust among direct users. Gradual implementation in phases, where the model does not hold authoritative power, facilitates organizational acceptance.

Based on insights from all sources, a step-wise plan for ML model implementation is proposed. The initial phase involves assembling the necessary infrastructure, data, and stakeholders, followed by creating and implementing a minimum viable product as a confirmation tool alongside the existing forecasting process. The minimal viable product is a functional model that provides usable accuracy and lists expected monthly sales based solely on historical data and observed fluctuations.

The second and final step focuses on refining the model and presenting its findings through a dashboard, incorporating information from other relevant sources to support the forecaster in making informed forecasts. This phase also enables improved communication with other relevant departments. Ultimately, the forecast incorporates real-time sales data for increased accuracy.

Feedback from Besi representatives indicates that this implementation approach is suitable for their business. The framework and steps presented in this study have been generalized to benefit other organizations, making an academic contribution in the field of ML-based demand forecasting. This contribution stems from the building upon the theory by Caniato in implementing a quantitative forecasting method. The new findings show that there are multiple ways to implement a quantitative method and that, according to this paper, the implementation is best cut up into two phases for smooth transition and maximum acceptance.

ACKNOWLEDGEMENTS

Dear reader,

This report is my master thesis with which I finalize my studies at the Delft University of Technology and my master Complex Systems Engineering & Management. During this thesis project I had the pleasure of working with many amazing people who helped make this research possible. I would like to express my gratitude to the people who helped me in this research.

First of all, my thanks goes out to all the people at Besi who I had the pleasure of working alongside for half a year. I am particularly grateful to Henk Jan Jonge Poerink, who gave me the opportunity to write my thesis at his office in Kuala Lumpur and assisted the whole process through great advice and coaching. Furthermore, many thanks go out to Andreas Schöpper, who is the specialist in forecasting and thus gave invaluable information on the subject.

Second, my thesis committee. Ron van Duin, Aaron Ding and Lóránt Tavasszy, thank you all for your insightful comments on my research and for steering me in the right direction. A special thanks goes out to Ron for always making time for me with any question I may have, be it content or process related. Without your guidance this project would have never been completed.

Lastly, I would like to thank the people from my personal circle. My housemates at the Verticas Residensi in Kuala Lumpur and the friends there who helped motivate and think outside the box. In addition to, my family and my girlfriend who always supported me and believed in me.

Olivier van Spengler

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

MLP Multi-Layer Perceptron

ML Machine Learning

BoM Bill of Materials

JIT Just In Time

Besi BE Semiconductor Industries

APac. Asia Pacific

AEX Amsterdam Exchange index

R&D Research & Development

SVR Support Vector Regression

ARIMA Autoregressive Integrated Moving Average

SVM Support Vector Machine

ANN Artificial Neural Network

AI Artificial Intelligence

ARIMAX Autoregressive Integrated Moving Average with Explanatory Variable

NN Neural Network

CoSEM Complex Systems Engineering & Management

BBP Besi Business Partner

FBA Forecast By Analogy

FSS Forecast Support System

DSR Design Science Research

1 INTRODUCTION

1.1 SUPPLY CHAIN

Effective supply chain management has become the core of practically all manufacturing companies. An efficient procurement plan and inventory management is essential to keep companies competitive. Deliveries should be met but with the smallest possible inventory levels for the most effective resource and financial planning. The just-in-time JIT principle defines these aspects as essential. The goal is to get the resources necessary for production in when production starts and only have the product finished when the client needs it.

Great supply chain management calls for planners that transcend company boundaries, involving multiple links of the supply chain Anderson et al. (2007). Suppliers need to contact their suppliers in time especially with some parts taking up to a year to produce. However, to be able to tell a supplier how much to produce in advance, an organisation needs to know the demand of its own customers in advance. This is where demand forecasting comes into play.

1.2 DEMAND FORECASTING

Accurately forecasting the demand of customers is essential in both long-term and short-term planning. The robustness and efficiency of the production process could be improved as the resources are better aligned with the demand Acar & Gardner Jr (2012) Ma et al. (2016). This means lower warehousing costs and less liability through on hand parts which may become obsolete. Furthermore, in addition to reduction of costs, the severity of possible bullwhip effects, failed deliveries and delays due to stockouts will be less frequently. This is another important aspect as the market in which Besi operates has a rather small base of customers. Customer retention is thus very important. Furthermore, for those customers a timely delivery is essential seeing as they are subject to the same rapid improvements in technology that forces them to stay up to date with the latest technologies.

Recent history has seen many forecasting techniques and models being developed. There are two approaches for demand forecasting, one is the forward approach and the other the backward approach Chase (2016). Forward demand forecasting looks at the potential demand in the long run, usually planning for multiple years ahead. It focuses on following the trend in the market to see where it is going. The backward approach utilises past and current data on demand to forecast what the demand will be in the short term. Backward forecasting is used to predict demand in the coming weeks or months, with some techniques becoming accurate enough to predict accurately up to a year.

The techniques include empirical, qualitative (customer survey, delphi method, expert panel, etc.) and quantitative (Naïve, seasonal, linear regression, support vector machines, artificial neural networks, etc.). The methods differ in accuracy, in long and short term, the type of data they work best with, how the data is used and their explainability. Therefore, there is no best technique but instead each type of demand forecast requires a different technique. Explainability is an important aspect of the forecasting technique Caniato et al. (2011). It must be made clear that no one and no model will ever have prophetic powers and predictions will always have a degree of uncertainty Makridakis et al. (2020). Besi, being a publicly traded company, therefore has to be careful when making predictions as it cannot rely completely on the models, in case it goes wrong. When it goes wrong even by the slightest of margins, Besi has to be able to explain to their stockholders why they made certain decisions.

1.3 BE SEMICONDUCTOR INDUSTRIES N.V. (BESI)

Besi is a Dutch company with its headquarters in Duiven, The Netherlands. Furthermore, the company is listed on the Euronext Amsterdam and included in the AEX. The AEX is a free float market capitalization weighted index that reflects the performance of the 25 largest and most actively traded shares listed on Euronext Amsterdam. Besi is a market leader in the "development, manufacturing, marketing, sales and service of semiconductor assembly equipment for the global semiconductor and electronics industries." Quote Besi (n.d.). Besi's customers are mainly leading manufacturers of microchips, electronics and industrial companies and device manufacturers. Besi APac. is the operations side of the company, sourcing and procuring parts, manufacturing and assembling the machines. 73% of Besi's total 2141 employees operate in Asia, primarily in Malaysia and China. However, the R&D and top management is still vested in Europe.

This study is conducted as a master thesis for Delft University of Technology and practical research for Besi. During the whole duration of the thesis project the writer worked in house at the Besi Apac. office in Malaysia. There, interesting conversations with employees brought to light the knowledge gap from the private sector. Furthermore, the use of their data, resources and knowledge greatly benefited the completion of this project.

The issue that Besi faces is that due to high demand volatility it is difficult to accurately forecast demand. This results in forecasts constantly changing meaning that orders have to be cancelled and hastily reordered when the forecast calls for it. This shows in the purchase order statistics. Around 45% of the orders are pull in meaning that the parts may arrive late, 12% are cancellations and 10% are push out meaning too much inventory was acquired. These purchase orders have to be made according to the forecast but add a lot of unnecessary work and costs in addition to potential conflicts with suppliers. The long lead times on certain parts and the fact that items become redundant quickly in the high end electronics market adds to the importance of accurate forecasts.

Machine learning techniques have shown to excel in forecasting for high volatility markets. Therefore, integrating such a technique into the existing forecasting process could greatly increase the hit rate and hopefully stabilise the forecast, thus lowering the pull in, push out and cancellation orders.

1.4 CONTENTS

The report starts off with a data description and literature review into the relevant subjects for this research. Topics such as purchase orders and total machine sales are tackled through interviews with employees of Besi and by diving into their own database. The literature review goes into the various types of forecasting, machine learning techniques, hybrid models and the usability of ML techniques. The chapter concludes with a problem statement and knowledge gap.

Chapter three details the design objective of this research and the path to getting there. It explains the actions that have to be performed at every step. A clear description of each of these steps is given and the subgoals that have to be met for going onto the next step. It concludes with a research flow diagram highlighting all these steps visually.

The fourth chapter, titled Design, starts the investigation into the current forecasting process and techniques. Furthermore, the limitations Besi experiences in its current forecasting format and implementing a machine learning technique are explored. The desires of the company and different actors within Besi are then explored. Lastly, the input and output of a machine learning forecasting model are explained. All these findings are then used in the following chapter to develop solutions.

The Development chapter does what the name suggests. This chapter is focused on creating the ideal forecasting process for Besi based on the data gathered in chapter 4. Important steps within the process are explained in more detail. The steps are: integration, monitoring, updating, forecasting and communicating. All of this results is an ideal forecasting process visualised in a framework in figure 5.1. The chapter also includes exploration of peripheral matters and the possible inclusion of a dashboard to aid the forecaster. The peripheral matters explored in more detail here are law and end user trust.

Chapter six is the Demonstration & Evaluation chapter. In this chapter all the findings and developments are shared with the vice president die attach of Besi, who is responsible for the sales forecasting of the die attach product group. His expertise gives interesting insights and challenges the findings on multiple facets. For example, the ideal process is not a goal he aspires to any time soon, a minimal viable product is what he desires. This also means that the development of a dashboard is very far away. The led to the development of a new process framework which is the first goal of implementation leading to the originally designed framework if the other has proven its worth. Lastly, a case study is performed with the same forecaster to test the bias there is towards a new ML model versus the current forecasting information sources. The Evaluation section evaluates the findings from chapter six. It looks deeper into what the findings mean for Besi but also for the wider academic community. This in terms of the implementation process and the results of the case study.

The Conclusion brings all the findings together and details the ideal course of action. Here, a generalised version of the implementation process, accompanied by multiple frameworks, is given so that other organisations can more easily relate the findings to their situation. In addition, concluding remarks on all other findings is given in this chapter.

Chapter eight is a critical reflection of the research process. Here the writer looks back at the complete process of doing the research and writing up the report. Subjects such as stakeholder satisfaction, communication, drawbacks and positives are discussed. The lessons learned from the whole experience are shared here.

2 PROBLEM IDENTIFICATION

As stated in the introduction, machine learning demand forecasting models have proven to be accurate and effective. They do, however, have their drawbacks. They could be hard to use and understand, plus, a lot of data is needed over a long period of time to make it effective.

Assignment from Besi: *Assist with the implementation of a previously proven, machine learning, forecasting tool.*

It is clear that a form of forecasting tool could work well at Besi according to Steenhuis (2017). However, implementing such a technique has proven to be difficult. This chapter takes that knowledge and investigates the sales data at Besi and highlights the state of the art of the relevant topics for this research. The topics include forecasting, time-series forecasting, machine learning, multi-layer perceptron MLP, hybrid models and usability. Each section also states the key search words and the database used for that topic, this is done to make the literature review replicable. Furthermore, the papers read for this literature review are listed schematically in the bibliography at the end of the paper. The chapter is concluded with a precise and concise problem statement and knowledge gap.

2.1 DATA DESCRIPTION

Key words: *multi-layer perceptron, high volatility demand forecasting, seasonal/-cyclical demand forecasting*
Sources: *BE Semiconductor Industries internal database, Google.scholar*

2.1.1 Purchase orders

Besi works with purchase orders for the acquisition of all the parts necessary to construct the machines. These purchase orders are based on the forecasted number of sales, usually nine to twelve months in advance. The orders are planned through backwards calculation. Meaning, if assembly of a machine starts on the first of January and the lead time of a part is two months then the part has to be ordered at the beginning of November. Therefore, if forecasts change in those last two months, either up or down, this leads to problems in the supply chain.

There are three things that can go wrong with a purchase order. The first is that there are not enough parts on hand and they have to be hastily bough, usually for a higher price. This is referred to as pull in. Another negative outcome is cancellation of an order. If the forecast changes to a lower amount while the parts have already been ordered then they have to be cancelled. Sometimes this results in still having to pay a fee to the supplier but in all cases it leads to problems for the supplier which should be avoided. The last issue is push out. Push out is when too many of the same part have been ordered and are now becoming redundant in the warehouse. These would have to be sold as a part if they cannot be used anymore, resulting in a loss.

Currently, of all purchase orders Besi experiences 45% pull in, 12% cancellation and 10% push out, as illustrated in figure 2.1. The desired thresholds are also illustrated in the figure. From the graph the evaluation is made that only the push out is within the desired threshold, and only marginally. The main cause for these high numbers is the constantly changing forecast.

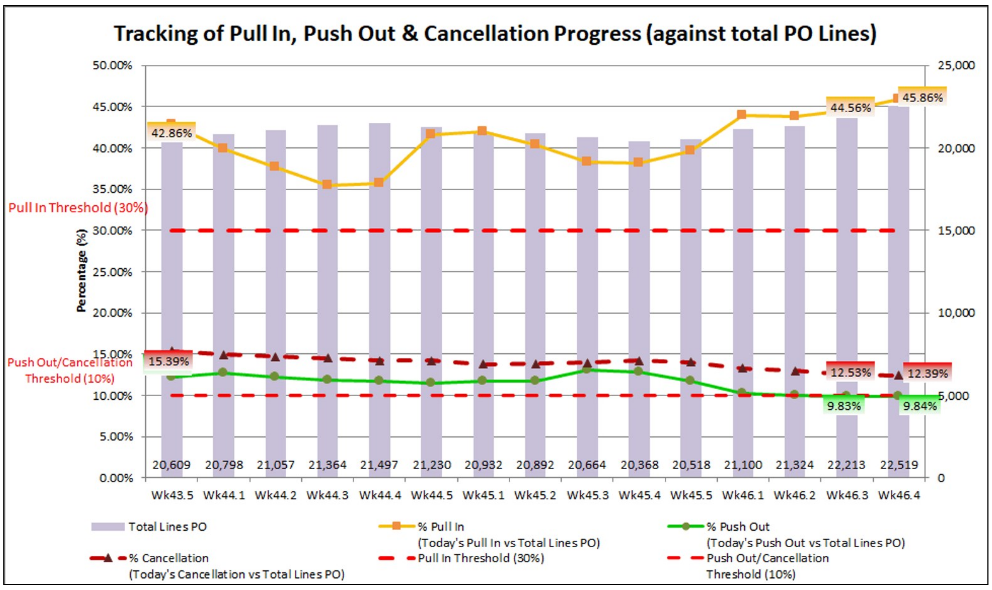


Figure 2.1: A graph showing the pull in, push out and cancellation progress against the total purchase order lines. Daily data from the end of week 43 to the fourth day of week 46.

Table 2.1 shows the weekly changing forecast for the next six months for one of the machines produced at Besi Apac. As is visible in the table, the forecast for the month of November was changed even in November. Further inspection of the table shows that December through to March also experienced great changes in forecasted sales.

2.1.2 Total machine sales

Initial literature research indicates that a multi-layer perceptron is the most effective algorithm for forecasting in high volatility markets. Nonetheless, analysing whether that is also the case for Besi is a wise decision. The unique seasonal and cyclical trends, shown by the total machine sales by Besi in figure 2.2, may be difficult to distinguish by machine learning algorithms.

Week	Oct'22	Nov'22	Dec'22	Jan'23	Feb'23	Mar'23	Apr'23	May'23	Jun'23
2022-39	8	34	24	20	13	45	45	68	56
2022-40	9	34	24	10	24	44	45	68	56
2022-41	8	32	27	7	29	42	45	68	56
2022-42	8	32	27	11	31	45	50	68	40
2022-43	8	32	27	9	31	47	50	68	40
2022-44	7	33	27	8	30	49	50	68	40
2022-45	7	33	27	8	30	49	50	68	40
2022-46	7	25	35	8	22	57	50	68	40

Table 2.1: A table that is communicated to supply chain managers within Besi every week about the forecasted sales for coming months.

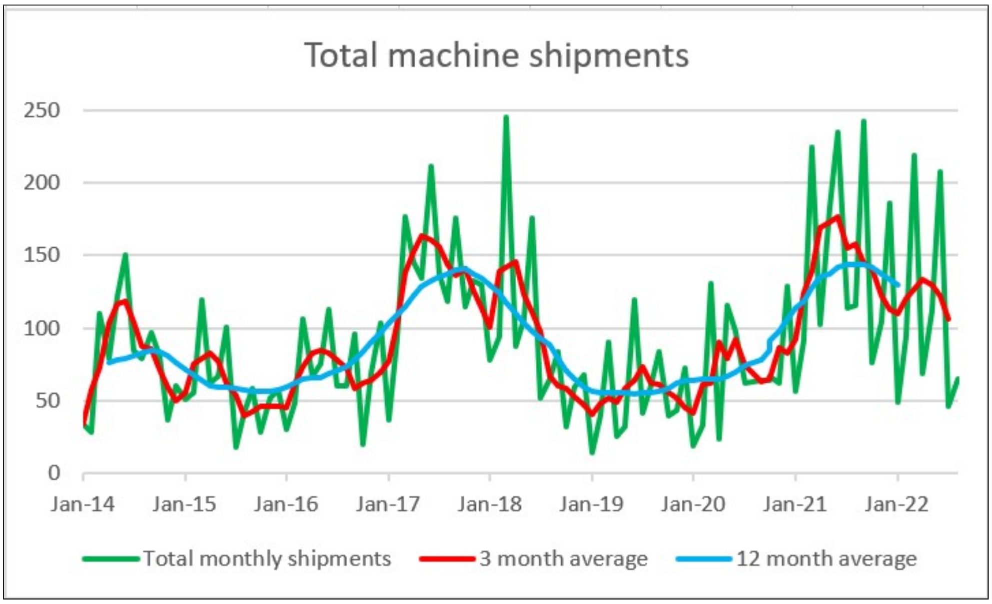


Figure 2.2: A graph showing the total monthly and three month average machine sales made by Besi Apac. since recording started.

Figure 2.2 shows the total monthly machine sales of Besi from 2013 till 2022. The moving averages are included to show the seasonal and cyclical patterns by averaging out the quarterly sales. There is a clear seasonal effect shown by the three month average, where every spring and summer there are more sales which cool down in autumn reaching its lowest point in mid winter. Furthermore, there is a cyclical pattern, highlighted by the twelve month average, that stands out with peaks from the beginning of 2017 to about halfway through 2018. Another cyclical top is evident from the start of 2021 and is estimated to end in the autumn of 2022; the data stops earlier but from current standing orders it is apparent that the peak is over.

Another interesting takeaway from the graph is that the seasonal and trend effects seem to grow in a multiplicative fashion. Multiplicative effects are defined by the seasonal element varying according to the underlying trend. In an additive model the seasonal event would remain consistent. This can be extracted from the graph by identifying that during the peaks of the 12 month average the three month average peaks and troughs are extremer than during the low points of the 12 month average.

This multiplicative growth in both effects is also visible in the time series data of product C in figure 2.3. Moroff compared various forecasting tech-

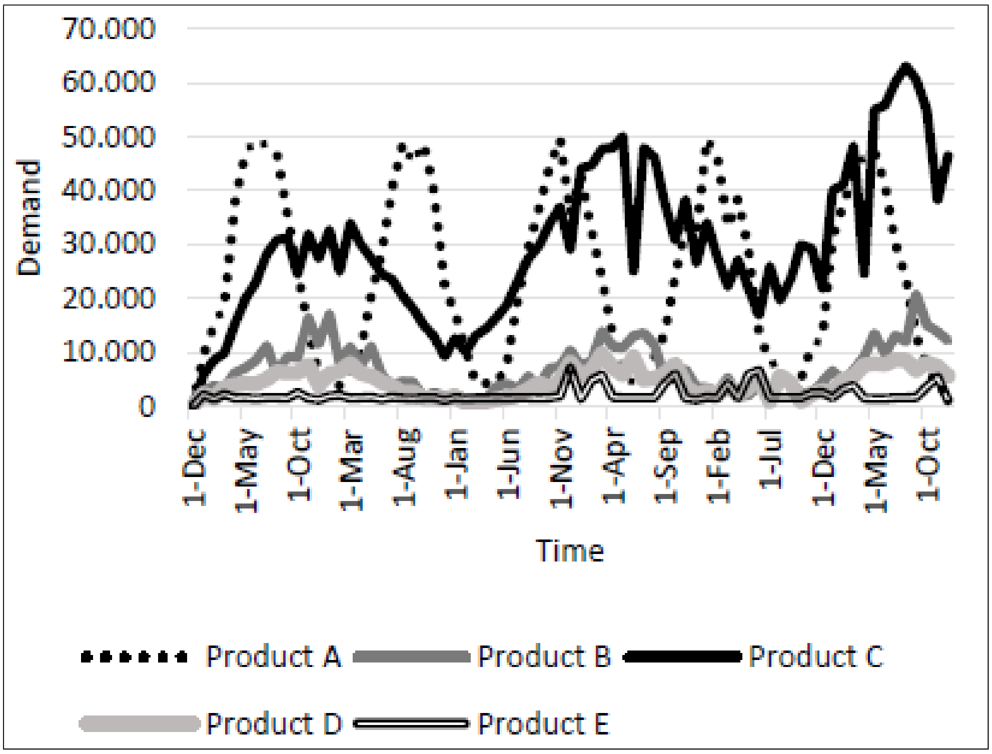


Figure 2.3: A graph illustrating the time series data used in a demand forecasting technique comparison study, extract from Moroff et al. (2021).

niques under different time series data patterns. Of all the datasets product C was the hardest to accurately forecast as is evident from table 2.2. The results show that for product C the best scoring technique, by some margin, is a multi-layer perceptron. This means that an MLP is best suited at accurately forecasting time series data with multiplicative effects in both seasonality and trend. Furthermore, prior research performed at Besi also showed that an MLP outperforms other techniques such as ARIMA, SVR and Naive Steenhuis (2017). Thus, a multi-layer perceptron is the most effective technique to apply to the case of Besi.

Prod.	Stat. Models		ML Models		DL Models	
	SARIMAX	ETS	RF	XGBoost	LSTM	MLP
A	4.341	6.786	10.817	9.706	3.932	6.172
B	8.109	3.219	8.777	8.712	4.658	3.656
C	14.237	13.513	25.477	25.978	24.539	11.591
D	1.909	1.879	3.902	3.983	2.633	2.313
E	1.010	1.208	1.439	729	1.034	954

Table 2.2: A table showing the root mean squared errors of the five techniques tested against the five datasets, extract from Moroff et al. (2021).

2.2 LITERATURE REVIEW

2.2.1 Forecasting

Key words: *Demand forecasting, supply chain forecasting*
Sources: *Google.scholar, repository.tudelft*

Forecasting is a tool used by organisations to be able to plan ahead in their business activities. All organisations need to plan their capacity management to be able to allocate their resources effectively. The just in time JIT inventory system method is an efficient way of cutting warehousing costs, however, it hinges on correct forecasting. If these forecasts fail it could cause shortages, downtime and panic buying.

Forecasting is the act of predicting what the demand will be based on relevant information. The forecasting methods can be categorised in two main categories. Qualitative and quantitative. Qualitative demand forecasting methods rely on opinions and instincts of sellers and experts. While on the other hand, quantitative methods focus on relationships between elements or patterns/pattern changes in historic data Chambers et al. (1971).

There are five main techniques that fall under these two categories as described by Hofmann Hofmann & Rutschmann (2018). Namely, Grassroots forecasting (sales force composite), Market research, Expert forecast estimation, Time-series forecast and Causal demand forecast. Each has their own benefits and drawbacks in various situations. Grassroots is where the organisation has a relationship with the buyer and therefore knows their intentions. Market research is when the whole market is analysed to get an understanding of the section of interest. Expert opinion is when an expert in the field of interest is invited to share their opinion on the matter using their prior knowledge. The following are the quantitative techniques. Time-series utilises historic sales figures and pattern recognition to predict demand. Causal demand applies knowledge about the distinctive factors that influence demand.

In this research the application of time-series forecasting will be analysed and how that fits into forecasting processes that could include multiple techniques.

2.2.2 Time-series forecasting

Key words: *quantitative demand forecasting techniques, time-series forecasting, machine learning time-series forecasting*
Sources: *Google.scholar, Elsevier*

Time-series forecasting is a forecasting technique bases its forecasts on one or multiple time series. A time series is group of observations made in a set sequence in time Chatfield (2000). The measurements can be made continuously or for a discrete set of time. Examples include, the length of a child measured each year at Christmas, the temperature every midnight in a fixed location or the total sales made of a certain product each month.

The large swathes of data that organisations have started collecting over recent years is one of the reasons why time-series forecasting has become possible. To be able to spot a reliable trend or pattern requires a lot of data. Furthermore, it has to be structured properly for analysis tools to be effective. Once that is all in order one or multiple of the many available time-series forecasting techniques can be applied.

Classical, or statistical, techniques include Autoregression (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving-Average (SARIMA). The machine learning ML techniques comprise of Support Vector Machine (SVM), Artificial Neural Network ANN, Long Short Term Memory (LSTM), Extreme Learning Machine (ELM) and Recurrent Neural Network (RNN).

Increasing research has been done into the application of machine learning (ML) for time series forecasting. However, research into the actual effectiveness of these models on improving supply chain performance is limited. Furthermore, little regard is taken for the inconsistencies of the effectiveness of these techniques and how that affects real life situations; plus, how the issues that arise from the incorrect forecasts can be mitigated Makridakis et al. (2020).

Various studies have been done into which is the most effective and accurate machine learning times series technique for certain situations. Carbonneau et al. (2007) compared ML techniques with classic demand forecasting techniques showing that ML techniques only marginally outperformed the classic methods. However, when applying a support vector machine (SVM) which was trained on multiple demand series they discovered that it had the highest accuracy of all techniques compared in the study.

Corbonneau continued his studies into the matter the next year, diving deeper into neural networks Carbonneau et al. (2008). This paper shows that while SVM had the highest accuracy on the training set, recurrent neural networks (RNN) showed the highest accuracy on the test set. This shows that the SVM SVM is limited in its ability for true generalization. Statistical analysis of the data showed that, although SVM and RNN performed better, there is no statistically significant difference in terms of accuracy between RNN, SVM and ANN. Moroff et al. (2021) shows that for some scenarios the more developed ML techniques do in fact outscore the statistical methods. It compared six different techniques on datasets that differed from each other by trend and seasonality proving that there is no one single model that works best. Artificial neural networks (ANN) do provide high level accuracy on time series forecasting when the time frame is lower Kochak & Sharma (2015). Furthermore, extreme learning machines (ELMs), which are essentially single hidden layer ANN, have been proven to drastically outperform ARIMA models, when applied in combination with another technique forming a hybrid model Chaudhuri & Alkan (2022).

2.2.3 Machine learning

Key words: *machine learning, machine learning demand forecasting, demand forecasting technique comparison*
Sources: *Google.scholar, Google*

Machine learning, or ML in short, is defined as the study of algorithms that learn from and improve with experience. It is a branch of computer science and artificial intelligence AI which focuses on using data and algorithms to mimic human learning. Through iterations the algorithm gradually improves its accuracy IBM (n.d.). It can distinguish patterns from data without the user having to define them a priori Murphy (2012).

In the growing field of data science, ML plays an important role. The techniques are effective in classification and making predictions through use of statistical methods. These insights could then be used for decision making processes within organisations IBM (n.d.).

The ML techniques include: Multi-Layer Perceptron MLP, Bayesian Neural Network (BNN), Radial Basis Functions (RBF), Generalized Regression Neural Networks (GRNN), also called kernel regression, Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), K-Nearest Neighbor regression (KNN), CART regression trees (CART), Support Vector Regression (SVR) and Gaussian Processes (GP) Brownlee (2019a). Of which the BNN and MLP score the best in the accuracy score Symmetric mean absolute percentage error (sMAPE) when testing for one-step-ahead forecasts Makridakis et al. (2018). Another comparative study titled A Comparison of Machine Learning and Classical Demand Forecasting Methods: A Case Study of Ecuadorian Textile Industry Lorente-Leyva et al. (2020) showed similar results. The study compared some of the most prominent forecasting techniques for demand forecasting in the Ecuadorian textile industry. Lorente showed that an MLP reaches the highest accuracy in all four test data sets, as depicted in figure 2.4.

Aside from the higher accuracy, another reason why a multi-layer perceptron would suite this research best is its simplicity. An MLP is the most basic artificial neural network. This makes it so that less time is spent on programming the algorithm and the focus instead can be directed on how the model is used.

The forecast of a machine learning technique is not perfect and will under perform in certain situations. Thus, trusting on it for the full 100% is in most cases not a good idea especially since they are black box techniques, meaning that the decision making process of the algorithm is untraceable and not explainable. Furthermore, the data put into the model has to be prepped. What this entails is that certain anomalies will have to be excluded from the training set. After the model prediction these anomalies, such as an exceptionally large order, have to added back in by human judgement. Therefore, combining a multi-layer perceptron with a qualitative technique is advised, thus creating a hybrid model.

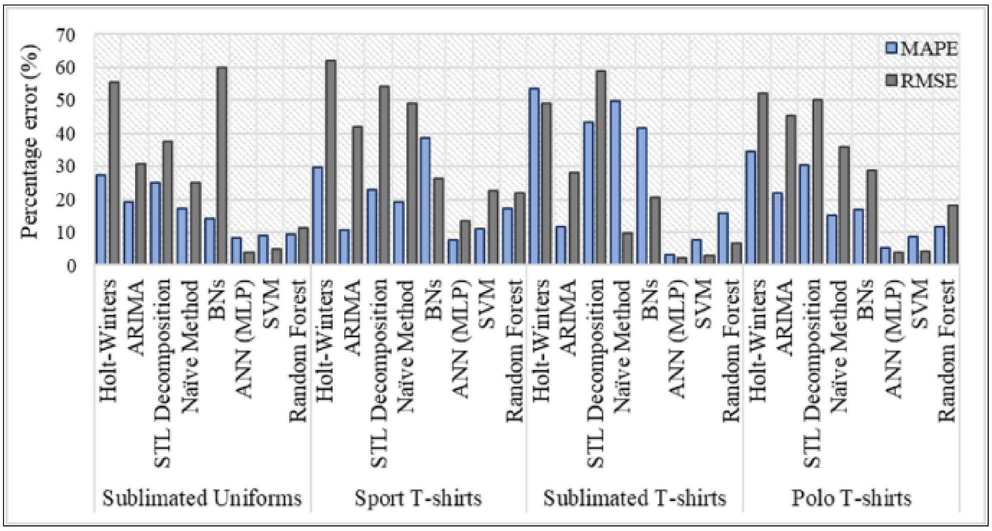


Figure 2.4: Results of comparison of prominent quantitative forecasting techniques Lorente-Leyva et al. (2020). The scores are given in mean absolute percentage error (MAPE) and root mean squared error (RMSE). Extract from Lorente-Leyva et al. (2020)

2.2.4 Multi-layer perceptron

Key words: *Multi-layer perceptron, neural network structure, neural network overfitting, activation function*
Sources: *Prior research performed by the writer, google.scholar*

Multi-layer perceptron MLP is the most widely used artificial neural network algorithm. The way an MLP works is it transforms constrained optimization problems into a series of unconstrained ones by incorporating the constraint functions into the objective function of the unconstrained problem. However, rather than solving the sequence of unconstrained problems through conventional unconstrained optimization methods we solve each optimization of the sequence by training an MLP network in a coupled neural network/objective-function representation J. Reifman & Feldman (2002). In other words, the goal of training a neural network is to find the weight w that minimizes the difference between predicted and desired outputs, i.e., the prediction error F. E. Reifman J. (2020).

A perceptron is the building block of an MLP. Its simplest form is illustrated in figure 2.5. The way it works is, the perceptron takes the input from the first layer and multiplies it by a certain weight w. After which it sums all the results (in this case only two) and adds a bias b. This means that the output h of this perceptron would be written as in 2.1.

$$h = x_1 * w_1 + x_2 * w_2 + b \tag{2.1}$$

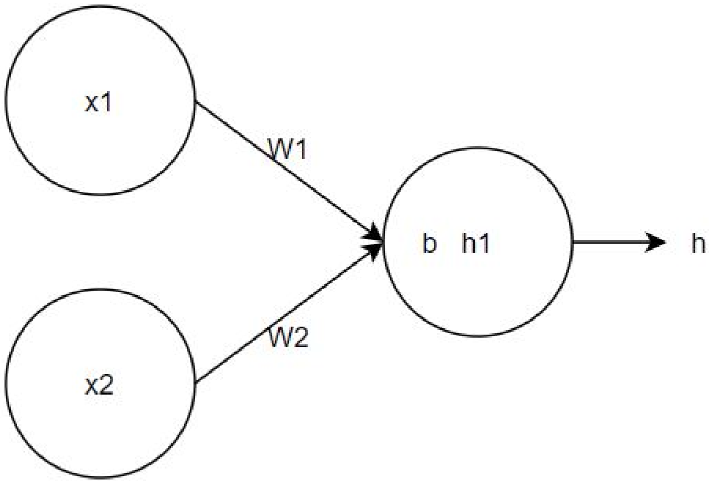


Figure 2.5: A single perceptron in its simplest form.

The general formula for a perceptron in a deep neural network is shown in equation 2.2. Where, h denotes the output into layer j. w_{ij} denotes the weight of node i in layer j and x_i represents the input from node i. The output, h, becomes the input for the next layer, shown as x, after the bias is added and an activation function is applied, which is explained further on in this section. Every layer in the neural network can be seen as a large matrix multiplication.

$$h_j = \sum_i w_{ij} x_i \tag{2.2}$$

If many of these are put into a layer and multiple layers are placed after another you get a deep neural network called a MultiLayer Perceptron which is depicted in Figure 2.6. Keep in mind that this is still a relatively simple version. The number of hidden layers can range from one to over five and the number of nodes per layer also ranges, all depending on the task at hand.

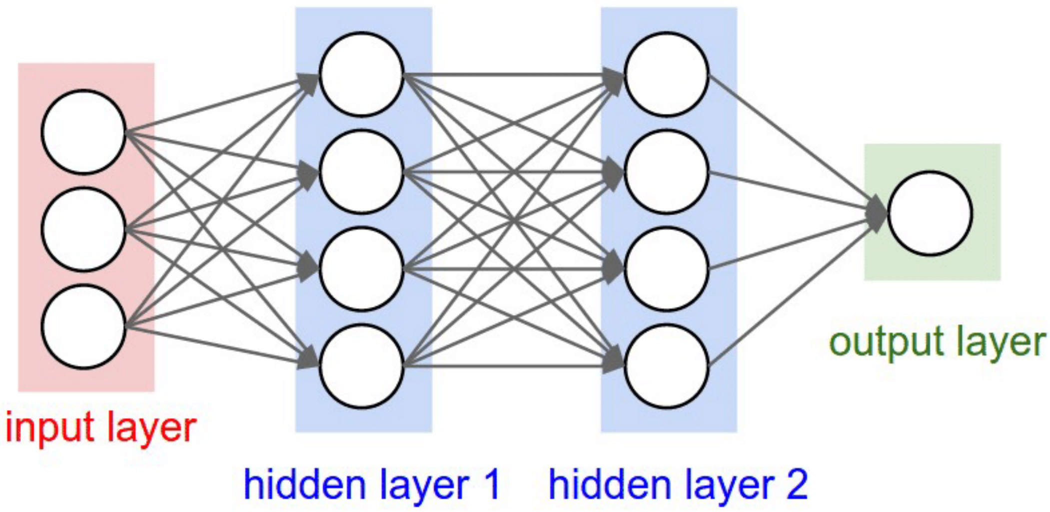


Figure 2.6: Simple deep neural network.

An MLP requires more than just perceptrons to operate properly. First of all, the data needs to be imported and split into training and validation/testing data. The training data, as the name would suggest, is used to train the algorithm, it uses these data points to figure out how to optimise its weights and biases. The validation data is used to check whether the algorithm is still improving or in fact over fitting to the training data. Overfitting is when the algorithm becomes too precise on the training data and therefor looses the ability to properly classify items it has not been trained on. This is best illustrated in Figure 2.7. It shows three scenarios where the middle one is the best. One may think that the graph on the right is good but remember that the two X's that it now identifies in the X category and logically outliers. Meaning that when utilising the algorithm on new data it would classify O's in that area as X's. In the last split, the testing data is used to test the completed algorithm to asses its accuracy.

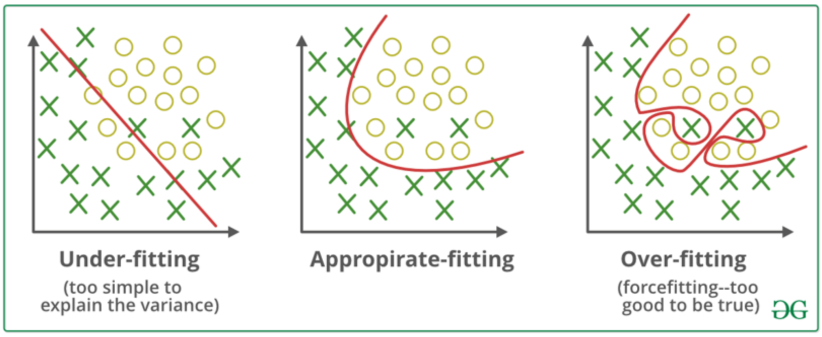


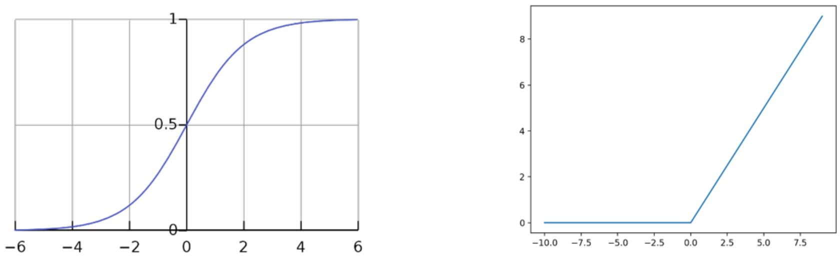
Figure 2.7: Three graphs illustrating the difference between underfitting, good fit and overfitting.

The setup of the model is the part that can differ the most. The setup consists of the number of layers, the number of nodes per layer, the type of activation function, whether it is fully connected or not, whether to add and how large the drop out should be, the type of loss-function, the optimizer, the metrics, etc. These are called hyperparameters and will all have to be optimised to be able to get the best results out of the algorithm.

The activation function is what is applied to the outcome of equation 2.2. An activation function is crucial for a neural network to operate because it introduces non-linearity. If this were not done then the sum of all the linear equations as in equation 2.2 through all the layers of the deep neural network, would still be a linear function. Which defeats the purpose of a neural network. The first activation function to be applied in neural networks was the Sigmoid function, described in equation **??** and depicted in figure 2.8a. Not only does it add non-linearity but it also normalises the output. All large numbers converge to one and large negative numbers to zero.

$$\sigma(x) = \frac{1}{1 - e^{-x}} \tag{2.3}$$

As more research was done into neural networks, new and more efficient activation functions were created. Rectified linear unit, ReLU in short, mathematically represented in equation 2.4, and shown in figure 2.8b is one such function.



(a) Sigmoid **(b)** ReLU

Figure 2.8: Sigmoid and ReLU activation functions

$$f(x) = max(0, x) \tag{2.4}$$

Fully connected is the term given to when all the nodes of one layer are connected to all the nodes of the following layer. Drop out is the act of deactivating a certain amount of nodes for every rerun. This is done to prevent overfitting as shown in figure 2.7. The goal of deactivating a node or set of nodes it to battle the over reliance of certain nodes. It makes every node train stronger separately.

For the last layer of a neural network a loss function or penalty function is applied. It is comparable to an activation function but only applied to the last node. This is done to determine the error in the prediction and then translate that error into better weights and biases in the nodes. One such penalty function is the Softmax function represented in 2.5.

$$\sigma(\vec{x})_i = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}} = \widehat{y}_i \tag{2.5}$$

This will result in the algorithm appointing the item with given input values to a class with a degree of certainty. This means that it will rarely know for 100 percent which class it actually is. Therefor, a loss function can be applied to the results to know how large the error is for each output. The loss function that is connected to the Softmax function, also known as the Softmax loss function or categorical cross entropy, which is described in equation 2.6. K in this case represents the output size, y_i the label of the class, so a one or a zero, and $\widehat{y}_i$ the probability, ranging between zero and one, that this class is in fact correctly classified. Another popular and simple loss function is the mean squared error loss function, depicted in 2.7.

$$Loss = \sum_{i=1}^{K} y_i \cdot log(\widehat{y}_i) \tag{2.6}$$

$$Loss = 1/N \sum_{i=1}^{N} (y_i - \widehat{y}_i)^2 \tag{2.7}$$

The error on the output then has to be translated through to all the individual nodes where every weight and bias has to be altered towards a better forecast. This is done by minimising the loss function on every run. This updating of the weights and biases is called supervised learning for the algorithm.

The general formula for calculating the error of an individual node is given in equation 2.8. It is important to note that this calculation is happening backwards, also known as back-propagation, here the output of the system is the new input. δ_j^h denotes the error term of node the node we want to calculate, w_{jk} are the weights j in layer k. δ_k^o is the preceding error and $f'(h_j)$ is the derivative of the activation function.

$$\delta_j^h = \sum w_{jk}\delta_k^o f'(h_j) \tag{2.8}$$

Before the algorithm can be activated/trained, the weights and biases of all the nodes have to get an initial value. It is important that these values are different from each other and random. This is because if they were for example all the same the algorithm would not optimize as it would change every weight by the same value.

The basis of training a neural network is gradient-descent, however, many different forms have been invented. How it works is: after assigning a class to the input and calculating the loss on every classification this then gets fed through the whole network backwards, this is called back propagation. While doing this it calculates the gradient of every node it passes though to determine how large its influence is on the output. When all of this has been calculated all the weights and biases can be altered based on the steepest gradient, therefor decreasing the calculated loss the most.

The weights w_{ij} and biases b_i are updated according to equations 2.9 and 2.10 respectively. η represents the decided upon learning rate and δ_j^h comes from the error term calculated in equation 2.8.

$$w_{ij} \rightarrow w_{ij} - \eta\delta_j^h x_i \tag{2.9}$$

$$b_i \rightarrow b_i - \eta\delta_j^h x_i \tag{2.10}$$

When all these steps are applied properly the result is a supervised learning neural network which is able to forecast data. However, issues can still arise, such as the earlier mentioned overfitting. To prevent overfitting, various techniques can be applied. One such technique, drop-out, has already been explained. Another is early stopping. Early stopping is the act of stopping the training of the algorithm at the ideal time. This moment is found by checking whether the loss is still decreasing on the validation set split from the total dataset in the beginning. If this loss starts increasing again it means the algorithm is starting to overfit, meaning the training should stop.

2.2.5 Hybrid models

Key words: *Hybrid demand forecasting, combining forecasting techniques*
Sources: *Google.scholar, Elsevier, Interview with the senior vice president global operations BESI and managing director Besi Apac.)*

Hybrid models are a combination of different forecasting techniques applied in unison. It goes further than an ensemble method which simple utilises the results of two or more different techniques. A hybrid model is a sophisticated model in which the various techniques intertwine providing input for each other, thus, reaching better accuracy and understanding in the results.

Most academic papers written on the subject of hybrid demand forecasting models apply a combination of a quantitative and qualitative forecasting techniques. The two approaches complement each others strengths and weaknesses and have a lot of overlap which helps decision making Feizabadi (2022). A combination of techniques can increase the accuracy, usability and explainability of the results greatly. Structures for varying hybrid models have been put forward by a limited number of papers. Caniato et al. (2011) created a structure, shown in figure 2.9; which incorporates the X-12 ARIMA ARIMA model in combination with judgement from the sales team. It shows how the two methods can best be intertwined instead of used separately for accuracy of the final forecast. Another structure for an hybrid model was proposed in Rožanec & Mladenić (2021). Rozanec's framework mainly focuses on increasing the explainability of the forecast.

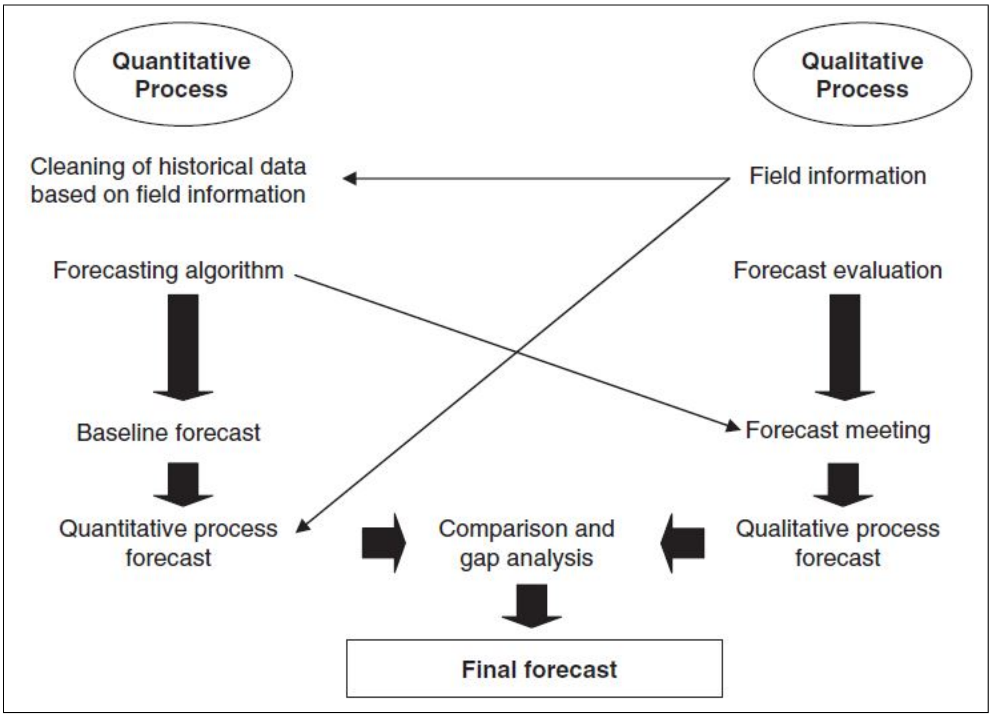


Figure 2.9: Structure for a hybrid model proposed by Caniato et al. (2011). Extract from Caniato et al. (2011)

Results from the study show that structured information from field research can significantly improve the quantitative forecast. Outliers that bias the quantitative forecast can be removed using historical data which is essential for understanding the demand that statistics cannot foresee Caniato et al. (2011). Aside from the greater accuracy, the hybrid model offered two more benefits. First, the forecasts became more consistent. Salespeople base their projections on a limited set of information and quantitative methods can be subject to unreliable data bias, combining the two clears out those inconsistencies. Second, the new method enabled the company from the case study to become more reactive. The new forecasting technique was easier and quicker than previous techniques so they could be done more often meaning that changes could be made more often Caniato et al. (2011).

Caniato also highlighted the importance of change management in incorporating the hybrid technique. The paper states that the implementation phase must be done diligently to gain acceptance within the organization Caniato et al. (2011). In the case of Besi it has already been said that the CEO is hesitant in relying on forecasting models, especially ones where it is hard to understand their working. Thereby accentuating the importance of explainability in the model.

Further research was done by Feizabadi Feizabadi (2022) into hybrid models. This hybrid model did not consist of a qualitative method. Instead, a combination of statistical and time series models was created, ARIMAX ARIMAX and neural network respectively. It proved that the combination of these two was also able to make up for the shortcomings of each individual model. The hybrid model addressed the vulnerability of excessive variance from the ML method and the bias issue if the statistical methods Makridakis et al. (2020).

The neural networks created a more smoothed prediction than the ARIMAX ARIMAX while the ARIMAX ARIMAX was better at predicting peaks. The paper also put forward a proposal for a new hybrid model, one which is based om multiple machine learning techniques blended together Feizabadi (2022). A combination of ML forecasting techniques could more effectively handle the model-, parameter- and data uncertainties Armstrong (1986). A model incorporating the findings from multiple separately tuned neural networks could in theory be a good demand forecasting model.

The senior vice president global operations BESI and managing director, senior vise president Besi Apac., stated that certain products are a good indication of the the trend of the market and others are better at predicting peaks. This is because of the nature of the machines. The ones showing the trend are the mainstream machines that form the building block of the total assembly line. Others are more specialist machines that usually precede the cyclical peaks in demand. A multivariate model incorporating the sales data from the different machines separately could make for an accurate forecast. This is known as a ensemble of experts coined by Slawek Smyl Smyl (2020).

One issue Henk Jan stated with such forecasting techniques is that it is too difficult to apply. To get a machine learning algorithm to work many steps have to undertaken, in a field that the planners have no expertise. Furthermore, because the workings of the algorithm are not completely understood, the trust is also missing. This is the main reason such a technique has not been applied yet even though prior research by a master student concluded that an ANN ANN can greatly increase the accuracy of short term forecasts Steenhuis (2017). The issue here is the usability of the algorithm.

2.2.6 Usability

Key words: *Machine learning, machine learning usability, explainability*
Sources: *Google, Interview with the senior vice president global operations BESI and managing director (Senior VP Besi)*

Forecasting is mostly done my marketing teams or logistics/supply chain managers. In general, these people do not posses the expertise to set up and use a machine learning algorithm for their forecasting task. This lack of knowledge creates a barrier to using such techniques.

For a long time the focus of research in the field of ML forecasting techniques has been in making the most accurate model. In recent years the explainability of AI, also referred to as XAI, has seen a growth in interest. The step that is clearly still missing in academics however, is the usability of the models. Large companies such as Amazon have the immense resources to hire the required expertise to apply effective ML techniques for forecasting, and are reaping the benefits Kit (2022). Smaller companies on the other hand do not experience the same luxury. For most organisations a simple multi-layer perceptron would be able to provide great results for their forecasting process, if used correctly.

The usability is not only about how to make and use the algorithm. It implies an understanding of what information is wanted from the results, what data is needed for that and how the data should be prepped, when the algorithm should be trained and retrained, how the results should be interpreted under differing circumstances, etc. A lot more comes to play when actually applying the technique.

2.2.7 Implementation

Key words: *Implementation techniques, AI implementation, implementation in businesses*
Sources: *Google, google.scholar*

As stated before there are many proven AI and ML techniques that could greatly increase the accuracy of forecasts. However, less research has been done into the actual implementation of such a technique and how that would go about. Caniato Caniato et al. (2011) dives into this issue by looking at what businesses need to properly implement and use a ML forecasting technique in their business.

In the paper from 2011, Caniato et al. (2011), Caniato explains that the best way for a business to utilise a quantitative technique is to integrate it into the existing qualitative technique. This way there is less resistance to the change and the forecasters can more easily adapt to the new way of working. The forecasting process and the organisation also need to be aligned so that a two-way flow of information from the centre to the periphery and vice versa is possible. This enables the integration of the new technique. Doing so would not only increase the accuracy but also increase the understanding within the organisation.

The implementation process must be carried out diligently and in multiple phases so as to ease it into the organisation. Caniato signs off with the statement that: *"According to this organizational perspective, a perfect and very accurate forecasting method might be useless if not implemented and used properly by people. On the contrary, less sophisticated methodologies might produce good performance when they are accepted and understood by the organization"* -quote from Caniato Caniato et al. (2011). Meaning that, equal, if not greater, attention should be given to the implementation process and peripheral matters that go along with it, as to the forecasting tool itself.

The implementation technique the Caniato uses in his work is the 'unfreeze, change, freeze' developed by Lewin Lewin et al. (1947). In Caniato's paper this technique worked effectively in implementing the quantitative tool into a qualitative forecasting system for a company operating in the cement industry.

2.3 PROBLEM STATEMENT

Effective demand forecasting in high demand volatility sectors is an exhausting task. Machine learning techniques could help ease this process, however, while the techniques them self are proven, the usability is still an issue for many organisations.

The proper way of incorporating the machine learning technique into their whole forecasting process is not understood. The results cannot be trusted blindly and having the correct input data is essential for the tool to work. Getting these things right is a tough task for planners and logistics managers who are generally not trained in the field of artificial intelligence.

2.4 KNOWLEDGE GAP

To the best of the writer's knowledge there is no literature exploring the implementation and usability of a machine learning ML method into a continual judgemental demand forecasting process in high demand volatility markets. There is no simplified framework that allows supply chain managers, lacking in artificial intelligence experience, to utilise machine learning ML.

3 RESEARCH FORMULATION

Literature research and interviews with Besi has allowed us to identify the knowledge gap stated in 2.4. This knowledge gap enables the formulation of an objective and research research steps which are mentioned in this chapter. Furthermore, this chapter will explain the scope, research approach, methods, activities, flow diagram and a time plan.

3.1 OBJECTIVE

To develop a demand forecasting framework that enables supply chain managers to properly integrate and use a machine learning, ML, model in their predominantly judgemental demand forecasting process.

3.2 RESEARCH APPROACH

This research will follow the structure of a design approach, more precisely the design science research (DSR) approach. This approach best fits the research as there is a gap in the knowledge of the functioning of a (complex) socio-technical system. Furthermore, a design in the form of a framework has to be made and proven to work.

The DSR approach is best explained in Peffers et al. (2007). The six steps in a DSR approach and their respective chapters in this research paper are: Identify the problem & motivate 1 & 2, Define objectives of a solution 3, Design & development 4 & 5, Demonstration & Evaluation 6 and Communication 7. Figure 3.1 is an extract from the paper by Peffers, it illustrates the flow of the steps and the iterative process of certain parts.

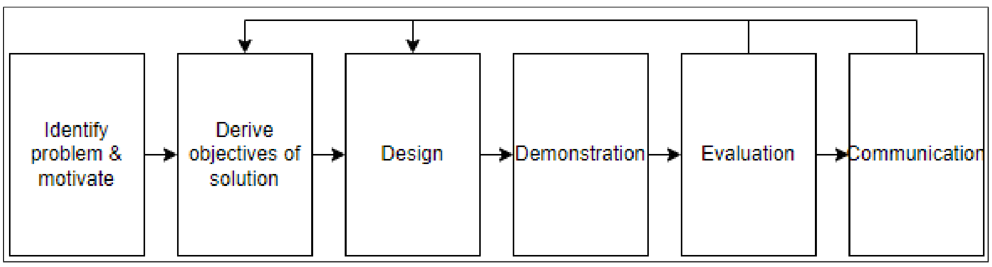


Figure 3.1: A schematic overview of the steps performed in a design research approach as suggested by Peffers et al. (2007).

This research leans more towards the practical side of design than the research as Besi is not only a test subject but also the sponsor of the project. Meaning they desire a more complete and specified product with less emphasis on the academic implications. Therefore, the work may be less structured in some ways, for example in the fact that there are more iterations between different stages in the DSR approach than only from evaluation back to design. Furthermore, design and development is split and demonstration is combined with evaluation as this fit the structure of the research and the practicality more.

3.3 SCOPE

In order to make the research project fit into the frame of a Complex Systems Engineering & Management master thesis it has to be scoped down. This scope will detail to the reader what exactly will be included and what not.

First of all, the goal is to develop a framework that incorporates a ML model into the demand forecasting method of organisations. However, no ML model will be developed. The literature review showed that an MLP is one of the best algorithms for demand forecasting in volatile markets. Furthermore, it is the simplest algorithm to code making it the ideal algorithm for this project.

Second, the framework will be developed mainly for BE Semiconductor Industries. The framework will be generalised so that it adds value to the broader academics. However, the main focus is Besi as the writer is under contract for them while writing the thesis. In addition, Besi supplies the data and information needed to complete and test this research.

As the framework is designed for Besi some aspects that could influence demand for other organisations are left out because they do not apply to Besi. For example Punia, Punia & Shankar (2022), explains the six influencing factors of demand for retail organisations which are: promotions, times series feature, economy indicator, weather, transaction details and store related. Of these six factors only three truly apply to Besi; these are promotions, time series feature and economy indicator. Time series is the focus of the MLP, the MLP will not incorporate promotions and economic indicators. However, the framework will incorporate those two factors as they do need to be accounted for. The other three variables, transaction details, store related and weather do not play a role in this research.

3.4 RESEARCH STEPS

Reaching the design objective will be done in separate steps. This is done to structure the research into the correct chronological order and to communicate the plan to all interested parties. All the steps are listed in this section. All steps have their own objective and deliverable. Section 3.5 illustrates in what chapter these steps are reported and how these steps flow from each other in 3.2.

3.4.1 Design

The first steps focus on identifying the system. The current forecasting process and technique within Besi is analysed in addition to answer to the question why an ANN ANN has not worked for them in the past. Furthermore, the desired insights that Besi wishes to get from the ML algorithm and the input and outputs are determined. Therefore, the five steps to be performed in this chapter are structured as follows.

Current forecasting process

First of all the current forecasting process has to be understood and visualised. This is necessary as it shows where the issues lie and how an ML model could fit into the whole process. The goal here is to visualise the process in a framework with the steps taken, what techniques are applied, what data is used and by whom.

For this step information is gathered from interviews with multiple employees at Besi. The planning department is split up into multiple layers. Starting in the various Besi business partners (BBPs) in Europe who make their forecasts and communicate those with Besi Apac. One on one interviews with these people gives answer to the question of how the forecasting process is structured and used within Besi.

To build up the framework, relevant literature is also applied. Data on essential steps in the field of demand forecasting will lay the foundation of knowledge that also helps in conducting the interviews. This information is retrieved from academic papers.

Current forecasting techniques

Aside from illustrating the flow of information through all the different chANN ANNels and layers of Besi, it is also worth taking a deeper look into the techniques that are applied for the actual forecasting. This section concludes with an overview of the techniques, in addition to, how those techniques best coincide with ML techniques through hybrid models must be understood.

Therefore, interviews with all the forecasters of the four BBPs and relevant literature research are the methods to be applied in this step. The four Besi business partners may all apply different techniques so interviewing all of them is advised. Accompanying literature that further details those techniques and their chemistry with ML techniques is gathered.

Limitations

In 2017 a master student of econometrics and operations research found that Besi could benefit greatly from applying a neural network forecasting technique Steenhuis (2017). The company was enthusiastic about the prospect but failed to put it to use. Examining what went wrong here indicates the issues surrounding usability of machine learning techniques in practice. This section should conclude with a list of issues that hindered the adoption of the artificial neural network and solutions to these problems.

Again interviews will play a critical role in reaching the goal of this step. The same people can be approached as in the first step, although the higher up in the planners pyramid the more relevant. In addition, the paper by a previous student Steenhuis (2017) serves to indicate what the intended idea of the technique is, who should use it and how. The paper may also indicate limitations of the research that led to it not being used.

Requirements

To develop a useful and applicable tool the requirements of the intended user have to be clear. The objective of this step is to put on paper distinct and attainable requirements of the intended users of the technique.

Interviews and literature review are again the research methods. The opinion of all the planners is valuable to this step as everyone will be affected by the implementation of the technique. Therefore, it is important to know what is desired in each layer of the planning process and how a ML tool can help achieve that. The literature is employed to answer the question whether the desires are realistic and attainable.

Input and output

A machine learning model requires a clear input and output to work effectively. What data is to be used, how is it structured how is the output given? These are the questions that are to be answered in this section. The goal is to clearly define the input and output that would enable a machine learning algorithm to operate optimally.

To answer these questions again interviews are conducted. From interviews and inspection of the available data, the relevant and feasible input can be understood. Furthermore, the desired output is already largely answered in the previous step. The output will be influence by the requirements, possibilities and the way it will be used.

3.4.2 Development

In the design science research structure this step is combined with the previous one in a single chapter titled Design & development. However, for the sake of structure the choice has been made to split these steps and divide it over two chapters.

The deliverable of this chapter is a working demand forecasting framework that incorporates a ML model which can then be demonstrated to Besi and evaluated. Most of the required information and design decisions will have been made in chapter 4. Therefore, the Development chapter is there to piece all the puzzle pieces together to form a working framework with all its peripheral matters.

The framework is based on the current process applied at Besi and how an machine learning model can best be incorporated into that process. It should be self explanatory and easy to understand for the intended users at Besi. This is achieved by visualising the steps and flows of information. In addition, the framework should be universal enough for it to be applicable to the greater academics.

Deliverables of earlier steps provide the bulk of information required to complete this one. Chapter 4 sketches a base framework and the requirements of the new framework to be designed. Literature again plays a role in this step as theory on hybrid models, specifically hybrid models incorporating a ML technique, helps design this framework.

3.4.3 Demonstration & Evaluation

The chapter demonstration, will detail the opinion of the forecaster(s) on the developments from chapter 5. The new technique, framework, implementation plan and any other additions found later in the research are presented to the forecaster(s). Since they are the experts in the field, and the one(s) that will use it once fully developed, their view on the developments is essential. It may lead to changes to better the design or serve as confirmation that the right conclusions have been drawn.

Furthermore, a small experiment with the forecaster(s) is held, with the goal to research the behaviour of the user of a ML forecasting tool under correct and incorrect output of the model and/or other sources. Results from this case study help the forecaster(s) better understand their own biases in using the ML model.

The case study is performed in collaboration with one or multiple forecasters of BESI. They are presented with model and sales team forecast just like they would in a real case scenario. Some are the exact same as the historical sales data from 2019 and others are fabricated to be on average 20% above or below that mark. Furthermore, the historical data of the years prior is supplied so that the whole forecasting process is simulated.

In total seven different scenarios are to be tested. The cases that are tested are a mix of correct, 20% over and 20% estimations from the product groups and the model. This gives nice cases but the ones where the results from both cases are the same are excluded, exccept where they are both correct. So there are the six scenarios where the where both sources are different and one where they are the same.

The results of the test are given and evaluated in chapter 6. Peffers describes the evaluation step as the measure of how well the designed artefact supports a solution to the defined problem Peffers et al. (2007). Thus, a look back at chapters 2 and 3 needs to be taken to see whether this is achieved.

Evaluate usability

Evaluation of the effectiveness of the framework is done by first analysing the accuracy of new forecasts and second by the ease of use for the intended users. If during the demonstration it becomes clear that some steps or parts of the framework are not understood then this has to be amended. The objective of this step is to list all the solvable and unsolvable issues surrounding the finalised framework.

The evaluation is done my analysing the data from the demonstration and comparing that to a base case. The base case is how the forecasting is currently run at Besi. The results from the test run should be discussed by a panel of experts consisting of logistics planners from Besi and experts from the Delft university of technology. Multiple full or partial simulations could be done and discussed depending on time and necessity.

3.4.4 Communication

The last activity to be performed according to Peffers Peffers et al. (2007) is communication. All the findings and developments are clearly documented in this research paper and its appendices. Furthermore, Besi will receive special instructions and possibly training on how to apply the framework. Lastly, final presentation will be given at either Besi Apac or the Delft university of technology where are the findings are presented and any misunderstandings can be further elaborated on.

3.5 RESEARCH FLOW DIAGRAM

To give a clear impression of the steps that need to be completed to reach the objective a research flow diagram has been constructed. Visible in figure 3.2. It shows the steps, chapters, sub-questions and type of data gathering. Furthermore, the deliverables of each stage of the research are mentioned.

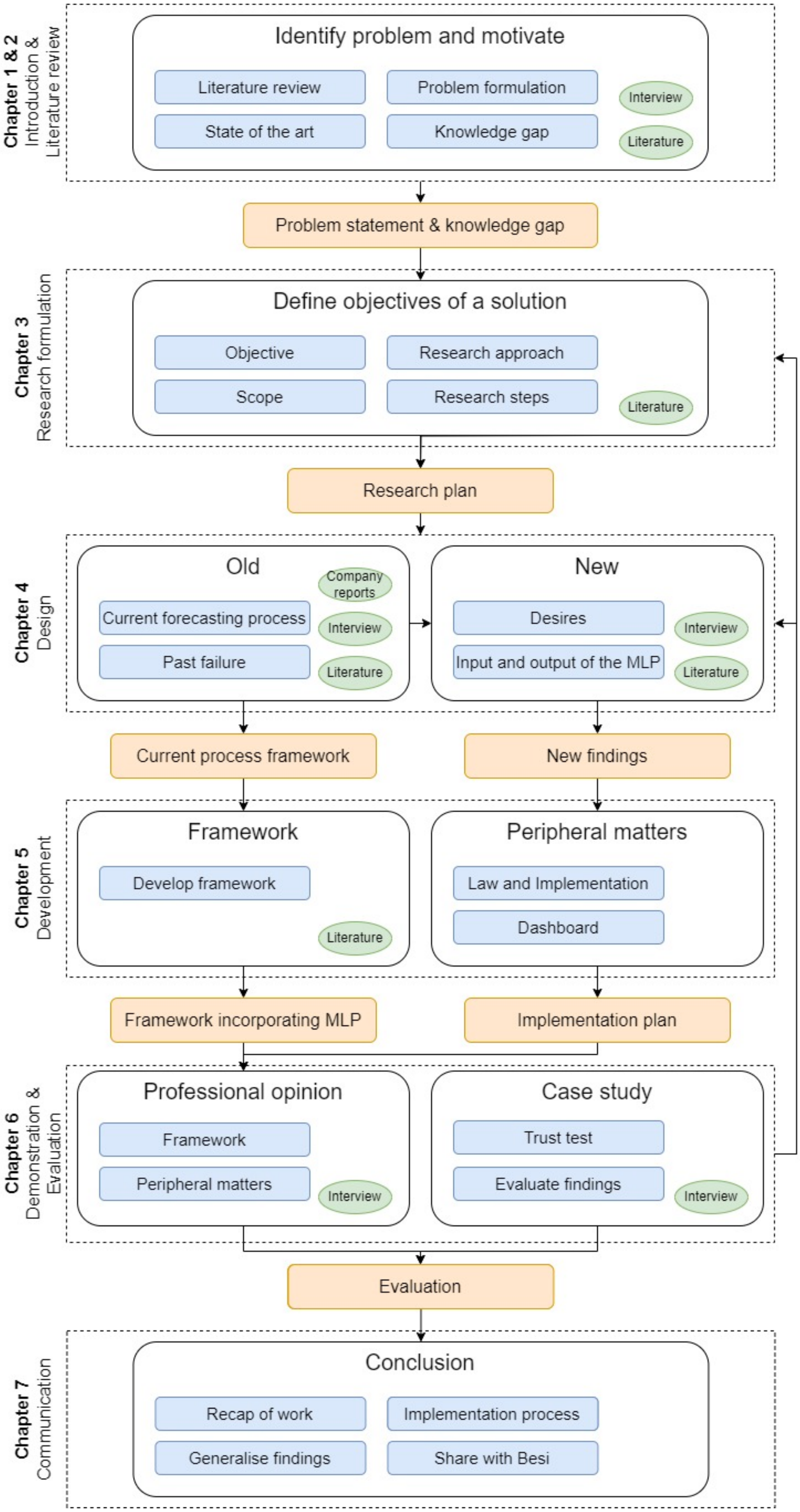


Figure 3.2: A flow diagram illustrating the flow of the research including the chapters, steps, sources and deliverables.

4 DESIGN

As the title would suggest, this chapter aims to understand the current process, and design a framework that would incorporate a machine learning ML model into that process. Therefore, section 4.1 looks at the current forecasting process at Besi and 4.2 looks at the current techniques applied. Not only that, but, the sections also dive into the alternatives and suggestions to improve are made based on literature. Furthermore, the limitations that Besi faces, section 4.3 and the requirements of the company, section 4.4, are discussed. The chapter is concluded by the detailing of the input and the output the machine learning model should have, section 4.5. The objective of this chapter is to set the knowledge base from which the framework can stem in chapter 5. Literature research and interviews with planners at Besi and experts from the Delft University of Technology are the sources of information in this chapter.

4.1 CURRENT FORECASTING PROCESS

To develop a forecasting framework with an ML forecaster integrated into it, first a base forecasting framework has to be developed. For this two relevant sources of information have to be consulted. First of all, the forecasters at Besi, section 4.1.1 and second, the literature, section 4.1.2. These sources combined give the knowledge necessary to develop a overview and understanding of the current structure. Resulting in the delineation of the steps that are included in the framework, made clear in table 4.1.

4.1.1 Interview research of current forecasting process

Prior research into the business structure and its processes has already been conducted by other students at Besi. However, additional interviews are to be conducted to confirm and build upon those findings.

Figure 4.1 is a flow diagram showing the physical, information and financial flows between departments within Besi, as constructed by a previous student. As is visible in the figure, the Besi business partners BBPs make the forecast and inform Besi Apac of that. There are four different Besi business partners, each with their own product line and headquarters. Each BBP is a separately registered company but all work towards the same goal. The BBPs are located in Austria, Switzerland, The Netherlands and Singapore. Their names also correspond to their location, so the one vested in The Netherlands is called Besi Netherlands, or BNL for short.

One of the issues for forecasting when there are four different entities is that they all make their own individual forecast. They have their own forecasting and sales departments for their own products. Although, all products are built in Malaysia and China who receive the four different forecasts and may not alter them as they are cost centres.

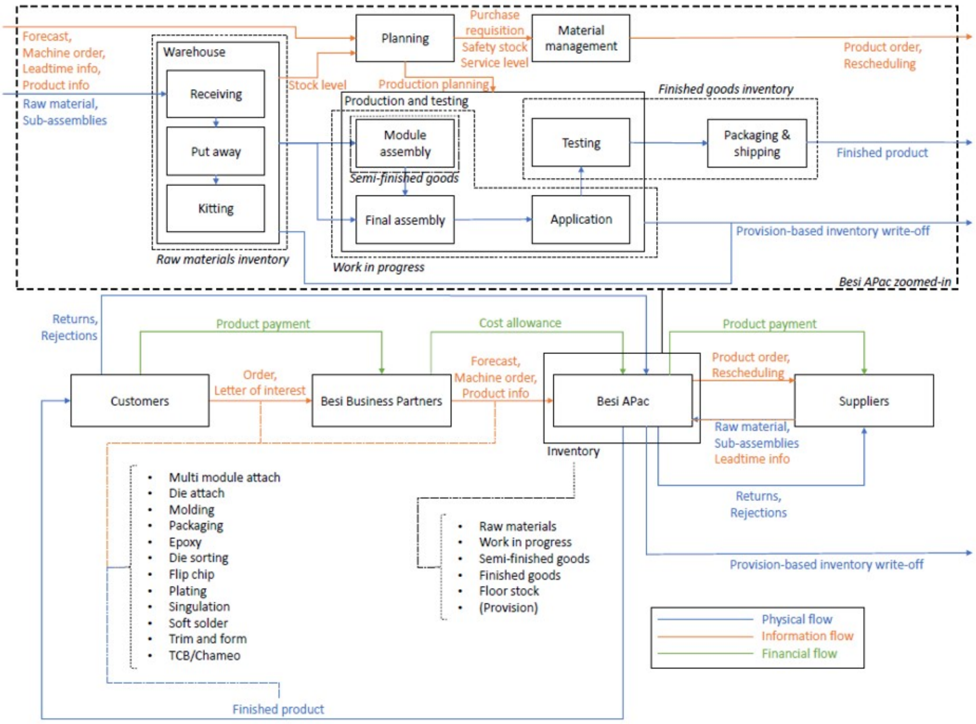


Figure 4.1: A flow diagram showing the physical, information and financial flows through the Besi company structure.

.

All sales teams, spread over the world, communicate with their clients and potential clients to the relevant product managers. Their prediction on the number of machines that these clients will buy is posted in a program called Radar. There it only shows the total amount of machines that need to be built of a certain machine family according to talks with customers. Next, The vice presidents of the various product groups take a look and make their forecast. The forecast is updated in Radar which is then a combination of the input from sales and the expected sales according to historical and market data. The consolidation of this data is done together with the product groups themselves. The data in Radar is also used to communicate with the shareholders, which emphasises the importance of transparency in the forecast. When the forecast is done it is updated in primary plan for further use (Appendix .1).

Managing planners and production planners retrieve the forecast from the primary plan and make their forecast for the different machines within the machine families. Their forecast is discussed with senior management for approval. When approved the information is passed on to Besi Apac. (Appendix .1).

The forecast is reviewed at different time intervals for different layers of the company. Besi Apac receives an update weekly but the vice presidents of

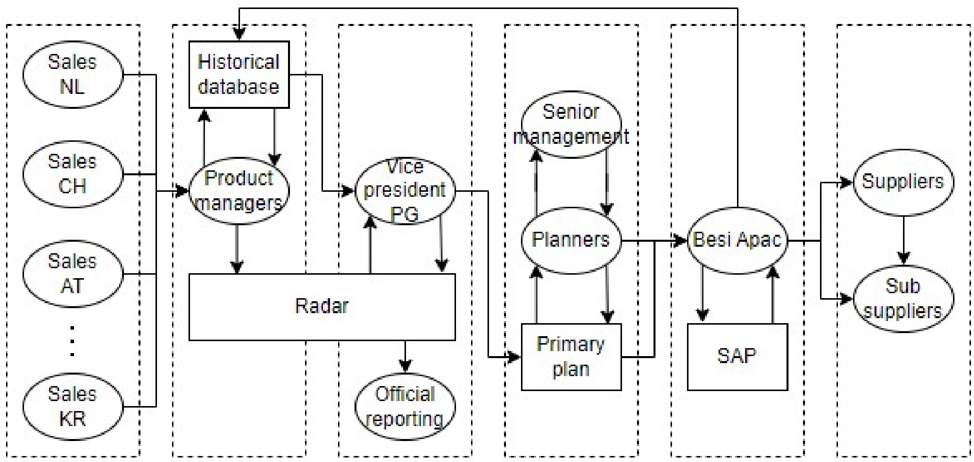


Figure 4.2: A flow diagram illustrating the flow of information in the sales forecast for machines per Besi business partner BBP.

.

the product groups only revise their forecast monthly. The planners update the forecast weekly to adjust for cancellations or quarterly adjustments. If Besi Apac does not agree with the forecast, they are not eligible to change the forecast, only to explain their concern to the business partners. The forecast reaches out between six to twelve months in advance. The same tactic is applied to next months forecast as the twelve month ahead forecast (Appendix .1).

The numbers that the planners receive from the forecasters consist of a confirmed part and a forecasted part. The confirmed orders are directly from the customer with a clear delivery date and also a clear BoM. However, even these orders can be cancelled or moved, making it hard to plan. The forecasted orders consist of ongoing talks with customers and the probability that these potential customers will indeed confirm their orders. The probability is referred to as the hit rate. This hit rate is based on gut feeling of the person talking to the customers. There is no systematic backing to this hit rate (Appendix .1).

There is a plan to switch from Radar and primary plan to a power BI system. This way the data that Besi Apac receives can be updated live and in a more clear structure. The power BI system will not replace the forecast made by the vice presidents, which is still necessary.

4.1.2 Literature research of current forecasting process

Chapter 2 showed that there is no universal way demand forecasting is done within organisations. However, there are distinctive traits that manifest in all processes that can be highlighted. Furthermore, the process of applying an artificial neural network for demand forecasting is discussed as this is the missing part of knowledge in the interviews.

There are multiple machine learning process descriptions such as CRISP-DM Shearer (2000), TDSP Severtson et al. (2017) and machine learning deployment workflow Ashmore et al. (2021). In this paper the focus is on the ML deployment workflow suggested by Ashmore, because it fits the structure at Besi well and is the only of the three for which the paper is available through the TU Delft.

The paper describes four main steps required for the deployment of a machine learning technique into business practices. These steps include: data management, model learning, model verification and model deployment Ashmore et al. (2021). These four steps are further expanded in actions illustrated in Ashmores paper.

For data management the actions are collection, preprocessing, augmentation and analysis. Data collection and preprocessing are already being performed at Besi to give structured, usable and accurate datasets to work with. Augmentation entails turning one datapoint into many, for reasons such as ethics or cost. This is not done at Besi and does not need to be done because of their low machine turn over, around 1500 in peak years, they can collect all the data they need. Analysis is an important task which is also diligently performed by each layer of the forecasting process as described in 4.2. After analysis the data can be augmented slightly by purposefully selecting relevant data and excluding outliers. When and how to include/exclude these data points is mentioned in chapter 5.

In the second step, model learning, the activities are model selection, training, hyperparameter selection and transfer learning. These are actions that are difficult to perform for a person without knowledge of AI and must therefore be made simple. The ideal model was researched in **??** and concluded to be a multi-layer perceptron, although this is based on literature and not testing so it could be altered. Training is the action of teaching the model to recognise patterns in the data so that it can predict the future. This will have to be done regularly to keep the algorithm accurate, how often must be described in this paper. Hyperparameter selection is best done by the AI specialist when creating the ML algorithm as it is important to get right for the accuracy and speed of the model. This action required trying multiple different setups of hyperparameters and seeing which one achieves the highest accuracy within a reasonable training time. Transfer learning means reusing old models to aid in speeding up the process of getting new models. This is a good option as the model will have to be trained more often but the model and hyperparameter selection do not have to be redone, thus, saving time and effort.

Model verification includes three activities: requirement encoding, test-based verification and formal verification. The first action, requirement encoding, requires both knowledge in coding and the domain in which research is to be done. Therefore, the AI expert must work together with the forecasters at Besi to produce the best tests to see whether the algorithm works as intended. Test-based verification is when test cases are used to asses the accuracy of the model. In the case for Besi historical data could best be used for this action. Although, future sales could also be used to test this if only historical data is used for the training. Formal verification is the last test to verify the model where irrefutable mathematical tests are performed. The results from these verification steps iterate back to the data management and model learning steps to improve upon the model.

The last step, model deployment, also has three main activities, namely: integration, monitoring and updating. Integration is entails implementing the ML tool into the wider forecasting process, which is what this paper is all about. How to alter the input data, integrate the output of the model in the wider forecast, protection of the system against wrong results, etc. All of this is further expanded on in chapter 5. Monitoring is also exactly what the word means, monitoring the model. This is done on various levels such as the input, the environment, the internals and the output. Constant monitoring is required for the safety, accuracy and relevancy of the system. This is especially important since the forecast is used to communicate with the shareholders. Updating is necessary to keep the model relevant. The first example is retraining of the model to keep up with the most recent data. But, other examples include applying the most recent techniques and models to increase accuracy. Since the field of machine learning is growing so fast new techniques and principles are invented in short succession, it is wise to update the system accordingly.

Khan et al. (2020) reports on effective demand forecasting model using business intelligence empowered with machine learning. 4.3 shows part of the paper where he illustrates the flow and transformation of information in the machine learning model. It shows the training of the algorithm, the prediction by the algorithm and the application of rules.

This part, figure 4.3 has to be infused into the process shown in 4.2. As is made clear from the interviews, and represented in figure 4.2, there are many layers in the forecast information flow. Thereby, highlighting the need for an accurate and stable forecast. Every layer decides its forecast based on the previous, so if the one before it is wrong or changes often then that negative effect only grows further down the chain.

Therefore, it is wise to tackle the forecasting issue as early in the process as possible. Sales is hard to change as that is very qualitatively done through direct talks with suppliers and is decentralised. The next layer is better to focus on. In the case of Besi Switzerland the vice president, vice president die attach, which makes finding and implementing a solution feasible. Furthermore, in this layer there is enough data to train a machine learning algorithm. In further layers, as the forecast expands to individual machine specs there is more randomness and less data making it harder to effectively apply a neural network.

The focus on the vice president die attach gives a more detailed picture of the interactions and data in/out flow that occurs when implementing the neural network.

First of all the data that is available. The input from the sales teams via Radar is one input, the others being historical data and information from the product groups. The interaction with the product managers gives an interesting dynamic where the vice president has to explain the forecast. Meaning there should be a degree of explainability and understandableness in the forecast from the NN. Lastly, the output is placed in primary plan which is updated weekly and gives the total machines expected to be sold in a certain month.

This links to the scheme designed by Khan in the following ways. The red object on the left side of the figure, 4.3, can be viewed as a combination of the program Radar and the historical database of Besi. The rule engine application is designed by the vice president based on the needs in terms

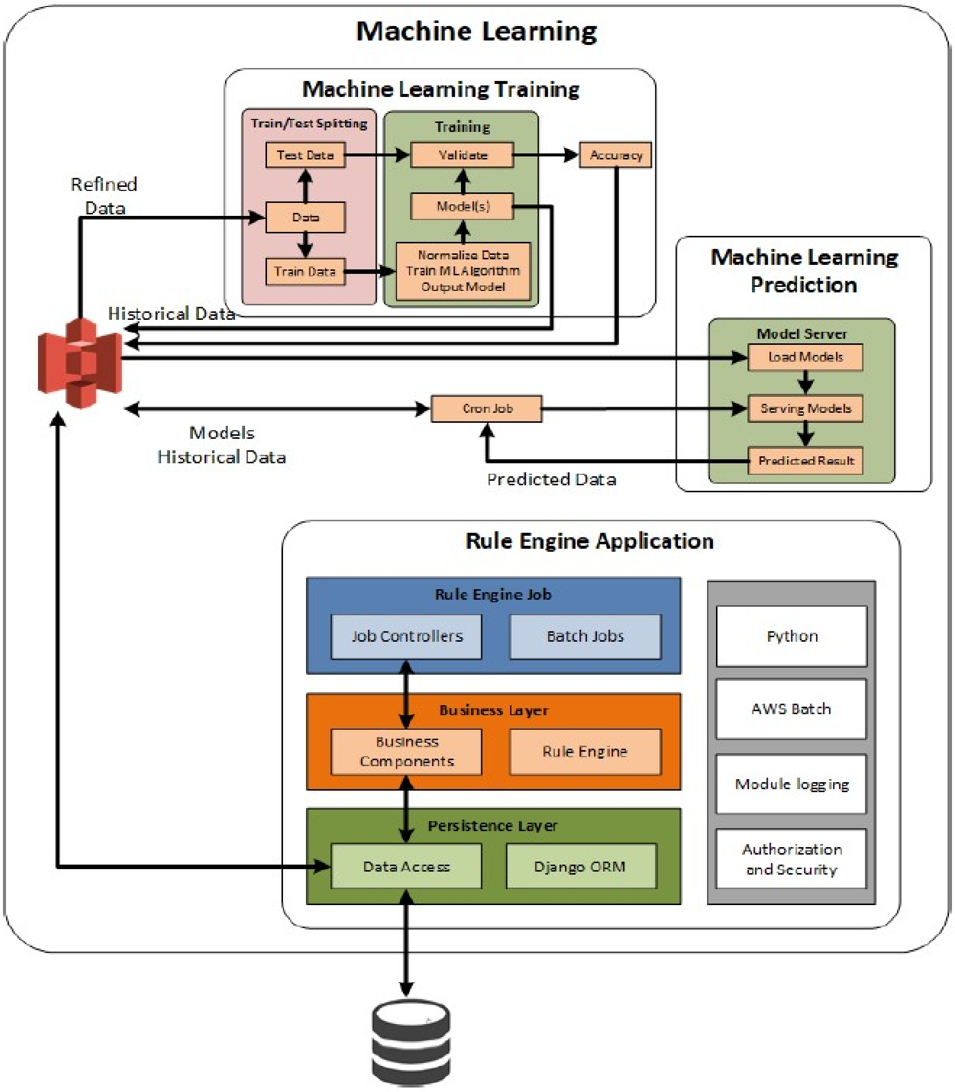


Figure 4.3: A flow diagram from a paper by Muhammad Adnan Khan illustrating the sources, steps and flow of information in the application of a neural network for demand forecasting Khan et al. (2020)

.

of accuracy, security, explainability, etc. This is the focus of further steps in the research. The machine learning prediction is the forecast made while the training of the algorithm is out of scope for this research.

4.1.3 Conclusion of current forecasting process

The forecasting process at Besi has many different layers with their own actors, data sources and actions. A machine learning forecasting model would work best for the third layer, from left to right, in figure 4.2. This is because it has the highest concentration of centralised relevant data, which makes the ML technique possible. Furthermore, this forecast is essential to get right, according to the managing planner and production planner respectively at Besi Switzerland (the people that have to work with this forecast in layer

4). This is because it is used as the most important source upon which the further layers base their forecast.

Furthermore, here there is one main actor that makes the forecast; thus, only one person who has to understand how to use the neural network model. This person, the vice president die attach, is interested in the project and excited to cooperate, which aids the research substantially.

The data for the forecast comes from three different sources: the historical database, radar (sales teams) and the product groups. This information is sorted by the vice president and he makes a judgement call on the final forecast which is sent to primary plan (Appendix .1). That is the scope of the forecasting process that this project will be focusing on.

Literature shows that there are some essential steps and actions that have to be performed for a machine learning technique to work effectively within a business. Some of these steps are already performed by Besi, some partially and others will be completely new. Table 4.1 illustrates the actions that require attention in the design from this paper. The actions are scored on the level at which they are already performed and whether the company can perform them itself.

Step	Action	Already present	Action needed	Outside help needed
Data management	Collecting	X		
	Preprosessing	X		
	Augmentation			
	Analysis	X		
Model learning	Model selection		X	X
	Training		X	X
	Hyperparameter selection		X	X
	Transfer learning		X	X
Model verification	Requirement encoding	X	X	X
	Test-based verification		X	X
	Formal verification	X	X	
Model deployment	Integration		X	X
	Monitoring		X	
	Updating		X	X

Table 4.1: Table showing the steps and actions as described in Ashmore et al. (2021) and how far Besi is in solving the individual actions.

.

4.2 CURRENT FORECASTING TECHNIQUES

Understanding and documenting the technique or techniques applied by the Besi business partners (BBPs) aids in the development of a hybrid technique incorporating an ML model. Interviews with relevant employees at Besi, shown in section 4.2.1, give insight into the current techniques applied. The literature research, section 4.2.2, takes the findings from the interviews are relays them with the existing literature. Thus, coming to a conclusion, section 4.2.3, on what the ideal new technique could be. Plus, diagram zoomed in on the part of the process in which the technique is applied is given, showing the current and the proposed information flows, figure 4.4.

4.2.1 Interview research of current forecasting techniques

The vice president of the die attach innovation department for Besi Switzerland responsible for the epoxy, multi module assembly, flip chip and SSI product lines, explained that the technique they apply for forecasting is almost purely based on judgement (Appendix .1). The information that the experts base their forecast on are the history, budget, talks with customers and predictions from the research and development team. History entails the trend and seasonal fluctuations of the market, the budget limits what is possible and the customer talks are the confirmed or partially confirmed orders. New machines coming from R&D, so without prior sales information, are forecasted mainly from the expectation of the R&D team but also on customer talks. The experts look at these bits of information on a certain day, and come to a conclusion on how many machines will have to produced in a certain month.

The way historical data is used in the forecast is simplistic. From previous sales, the seasonal trend can be seen clearly, as discussed in 2.1. The forecasters know that in general there will be more sales in summer than in winter. This knowledge is used to crank up the expected sales in the summer months even if the sales team does not show that yet. The seasonal effect is known and used to a certain degree, but the cyclical effect is not. Interviews highlighted the fact that Besi has been too slow to respond when needing to up- or downscale. Especially in times like the current when there is a downturn on the seasonal and cyclical effect at the same time. Showing once more that an unbiased forecasting model based on historical data as input can be of great effect in the overall forecast.

In the end the forecast is almost purely a judgement call. As is visible in figure 4.1 there are many layers of judgement calls overlapping each other. Meaning that more and more bias can creep into the forecast.

Another (sub-optimal) technique Besi applies throughout its different forecasting layers is a hit rate (Appendix .1). A machine usually consists of 50-70 unique parts. However, Besi offers its customers over 200 different options. So long as there is no confirmed BoM, the demand for these options has to be estimated. This is done based on historical information and gut feeling. It is a point of hindrance for many within the organisation because the hit rates are often wrong. This is due to the rapidly changing demand for new products and the limited data for most parts. Due to this limited data though, this is not something that machine learning techniques can help with. What will help is if the first forecast made by the sales team is more stable and accurate because that will allow further supply chain managers and planners to focus on improving their hit rates, instead of playing catch up to the ever changing forecast they receive.

4.2.2 Literature research of current forecasting techniques

The only technique applied by Besi right now is judgemental. Even in cases where an organisation utilises a quantitative method, a judgement method is usually still applied over it. As will also be the case at Besi according to the vice president die attach and the senior vice president global operations BESI and managing director (Appendix .1).

Judgemental forecasts have many benefits over statistical methods, but, also drawbacks. First, the benefits, judgemental techniques perform better when there is less data available and it can take a lot more data into consideration which would be difficult for a statistical method to structure Caniato et al. (2011). Drawbacks are that it could take longer and it is generally subject to bias Caniato et al. (2011). This bias aspect can hurt a company, Besi has proven. One solution is to combine the judgement with statistical methods.

There are multiple ways of combining qualitative and quantitative methods. Armstrong & Collopy (1998) mentions three such ways. The first is to have a human judge the data that is entered into the quantitative model. This means, whether it would be wise to include relevant datapoints or exclude a certain customer due to their unpredictability. Second, use judgement to decide upon the best statistical method to apply. Third, the incorporation of domain knowledge into the forecast such as the course of the economy. All three human adjustments are interesting to take into consideration when designing the hybrid framework. The second step has already been performed in the literature review, namely the conclusion that a multi-layer perceptron would be the best technique. Step one and three will have to be performed for every individual forecast made. However, general rules for these steps will be determined in this paper. Step one is discussed in section 4.5.

There are multiple ways of integrating a qualitative and quantitative method. There are certain scenarios where a forecast will benefit greatly from judgemental adjustment, which is one of the ways to combine techniques. For example, for forecasts with long horizons and for products with high demand volatility Arvan et al. (2019). Generally, anything that is not incorporated into the statistical model has to be done by human judgement.

Another method is to combine forecasts. The simplest form is to have a statistical forecast and a judgemental forecast and assign them equal weights and see the outcome. Through testing the weights can be adjusted to make the forecast more accurate.

Arvan et al. (2019) also mentions other, less known methods of combining techniques. Judgemental bootstrapping for example. This method is used to cancel out inconsistencies in the judgemental forecast, however this is not interesting for Besi as it does not work well in unpredictable markets. Another method is rule-based forecasting (RBF), however this one also does not make the mark as it requires a strong trend in the data. Decomposition method decomposes the forecast into smaller parts of which some are performed by statistical techniques and the others by judgement. This technique could be interesting to partly introduce as dividing the input data from large customers into separate parts, as explained by Armstrong (1986), could lead to higher accuracy.

The last two methods mentioned in the paper by Arvan Arvan et al. (2019) are group forecasting and forecast by analogy (FBA). These are interesting for this research. Group forecasting is already done by the different layers constantly revising the forecast given to them. FBA, is a technique stemming from machine learning. It is a technique that works best for predicting demand of new products and products in the face of promotion. What the model does is it assumes that two different events share the same behavioural pattern. FBA can best be used in combination with the other integration methods. This has great potential for new product launches in the future. For now though it is not as interesting. This is because first a model must be adopted, this is most easily attainable by making a simple model which is accurate and easy to use. Through the base of such a model, further expansions can be made such as new product forecasting.

Forecasting support system (FSS) is defined as a framework, software or structured procedure to apply both expert judgement and statistical techniques. It integrates historical and contextual data to provide the forecaster with a valuable support system to aid in creating more accurate forecasts Arvan et al. (2019). Not only that but an FSS could also indicate uncertainties and help in explainability as multiple techniques are applied that either confirm or challenge each others predictions. Arvan et al. (2019) also indicates that an FSS is a good structure for incorporating multiple integration approaches. Therefore, it goes well with the FBA which also benefits from applying multiple methods.

4.2.3 Conclusion of current forecasting technique

The interviews and accompanying literature review describe the ideal combination of a judgemental and machine learning technique, usable by a product group vice president at Besi, or any other organisation. This ideal setup is through a forecasting support system (FSS) and may be based on the forecasting by analogy (FBA) method. However, case study tests must be performed to justifiably conclude this statement and find the best setup. The FBA is useful for new product launches but the first goal is to get a simple model to be accepted and used by the company so the addition of new product launch forecasting would only come into play later.

This way the ML forecast can be integrated into the current forecasting technique which is mainly based on judgement. Of the three sources of information: historical database, Radar (sales team) and the product groups, two are presented quantitatively and can be used to train an artificial neural network. These are the historical sales figures and the data in Radar. It would be wise to start off with only one of the two as input for the neural network and the choice should fall onto the historical data for that. This is because the information in Radar is already calculated and structured ready to use by a forecaster. The historical data is only the knowledge that there is a seasonal effect, no more detail than that.

The focus of this research is on the forecast made by the vice presidents of the product groups. Their inflow of information is illustrated in figure 4.4 as it currently and what the proposed systems would potentially look like. Included is a possible link between Radar and the neural network as well; however, the value and possibility of this has to be further researched.

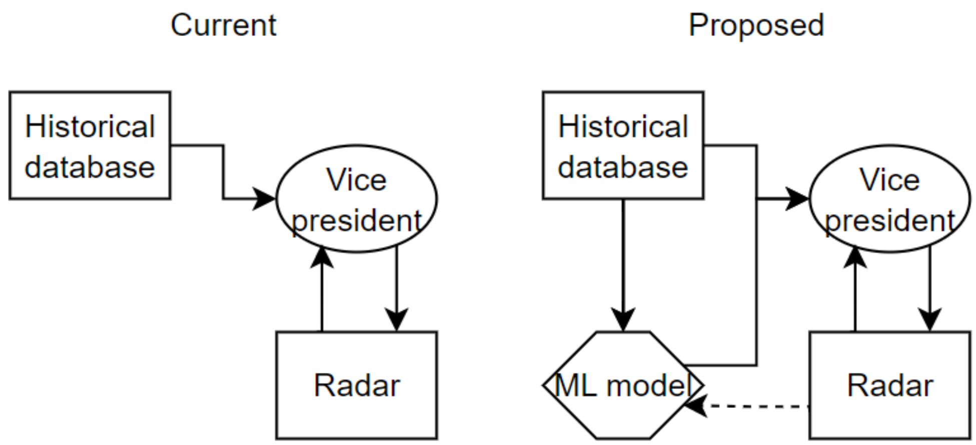


Figure 4.4: A zoomed in figure of the third and partially of the second step in figure 4.1. This figure illustrates the current information input for the forecast and the proposed implementation of the neural network in two phases.

4.3 LIMITATIONS

The research carried out by a previous student for Besi proved that an artificial neural network could serve as a strong tool in forecasting. However, it was never implemented. There are multiple possible reasons for this. This section investigates the actual reasons why the company did not implement the technique through interviews, section 4.3.1, and if/how these limitations can be overcome. Literature also aids in explaining the limitations of Besi and the more general issues for companies facing forecasting and forecasting with ML algorithms, section 4.3.2. In addition, other challenges that the company is facing in their forecasting process is analysed. The section concludes with a list of concerns, list 4.3.3, based on the research by Ashmore Ashmore et al. (2021), section 4.3.3.

4.3.1 Interview research of limitation

First of all, it is important to note that the previous research was carried out at Besi Apac. However, Besi Apac. does not perform the forecasts. So, the knowledge that the previous research provided was of no use to them. The findings were shared to others within the company who could have done something with it, but they never did. The vice president die attach innovation at Besi, did not know of the findings until the writer brought it to his attention (Appendix .1). This is the main reason that nothing was ever done with the knowledge that a neural network could aid in forecasting for Besi. Another reason, as explained earlier, is that the people who knew about it were not knowledgeable enough on the subject to be able to apply it.

Besi experiences more struggles when forecasting. One is that the customers cancel orders, another is that the hit rate of potential customers can be far off and yet another is that everything is based too much on human input. The option for cancellations for customers is a business choice made to attract clients, so that cannot be changed in the forecast. The hit rate is a completely different input into the forecasting process. Experiments could be done to increase the accuracy of the hit rate, maybe even using neural networks, but that does not fall within the scope of the research. The fact that it is almost purely based on human input, however, has influence on the forecast and can be mitigated.

Human input adds great biases in the forecast (Appendix .1). This is one issue delineated by the vice president die attach. The human factor is important, especially in low volume markets with high costs. However, it is slow to react to changes in demand. People tend to be optimistic when things are going well and pessimistic if the opposite is true Eroglu & Croxton (2010) Sharot (2011). This means that forecasters will ride the high of good times deep into the bad times, even when the signs all show that it is going down. The same is true for Besi according to the vice president die attach (Appendix .1).

4.3.2 Literature research of limitations

Common issues reported in the literature and online point towards the lack of knowledge as the main issue for organisation not implementing a machine learning forecasting tool. Other common issues are that training of the algorithm is too much of an exhausting task or there is no correct data available or not structured properly. Finding the right type of algorithm and configuring it correctly with the right training set are more problems that are brought forward by the literature.

This lack of knowledge spreads out into multiple issues. First, people do not trust such a system, then they do not know how to implement it and lastly there is not enough knowledge on AI to utilise it.

Building upon the ML deployment framework of Ashmore Ashmore et al. (2021), mentioned in 4.1.2, Paleyes case study survey Paleyes et al. (2020) addresses the issues regarding the steps in Ashmore's framework. These considerations, issues and concerns are illustrated in table 4.2. Paleyes also included an extra deployment stage. Namely the cross-cutting aspects, which are an interesting topic to take into consideration too. The following part is primarily based on the papers of Paleyes Paleyes et al. (2020) and Ashmore Ashmore et al. (2021).

When looking back at table 4.1 it is clear that some of these issues seen in the literature already do not matter for Besi. This encompasses everything in the data management section. All those steps are already correctly performed so the issues are nonexistent.

Other issues, however, do also affect Besi. From the stage model learning, the step, model selection is also done. Model selection has been performed in chapter 2 and that is where the direct involvement of the writer ends for this deployment stage. Although the other steps are to be performed by an external party with expertise in AI, it is interesting to think about these concerns. The computational cost may not be any issue for a high tech company such as Besi but it does add extra time to the training. If the training is to be done every week then preferably the training should not take too long. High computing cost also results in a larger negative impact on the environment which Besi is trying to subdue at the moment. So keeping the computational costs minimal by for example making a shallow (not many layers) neural network is a wise choice. Privacy aware training is another really important aspect for Besi. Besi has some of the worlds largest customers. If their data were to leak to the public this would have grave implications for that company and for Besi. Therefore, it is essential that the data is handled securely. This should not be an issue though. The input data for the model can be anonymous as it does not add much, if any accuracy to the outcome of the model.

Hyperparameter selection is another step to be performed by an external. This ties into the earlier discussed computational cost and environmental impact. As mentioned a shallow neural network would aid those two issues. The choice for the number of layers is a hyperparameter. Other hyperparameters include: the number of nodes, the cost function the interconnections, etc. As described in 2.2.4. Selecting the right setup of hyperparameters lowers the computational cost and time while simultaneously increasing the accuracy. Hardware aware optimisation entails that the computational cost should not exceed what is possible with the given computer, which from interviews seems to not be an issue.

Deployment Stage	Deployment Step	Considerations, Issues, and Concerns
Data management	Data collection	Data discovery
	Data preprocessing	Data dispersion Data cleaning
	Data augmentation	Labeling of large volumes of data Access to experts Lack of high-variance data
	Data analysis	Data profiling
Model learning	Model selection	Model complexity Resource-constrained environments Interpretability of the model
	Training	Computational cost Environmental impact Privacy-aware training
	Hyper-parameter selection	Resource-heavy techniques Unknown search space Hardware-aware optimization
Model verification	Requirement encoding	Performance metrics Business-driven metrics
	Formal verification	Regulatory frameworks
	Test-based verification	Simulation-based testing Data validation routines Edge case testing
Model deployment	Integration	Operational support Reuse of code and models Software engineering anti-patterns Mixed team dynamics
	Monitoring	Feedback loops Outlier detection Custom design tooling
	Updating	Concept drift Continuous delivery
Cross-cutting aspects	Ethics	Aggravation of biases Fairness and accountability Authorship Decision-making
	Law	Country-level regulations Abiding by existing legislation Focus on technical solution only
	End-users' trust	Involvement of end-users User experience Explainability score
	Security	Data poisoning Model stealing Model inversion

Table 4.2: A table showing the considerations, issues and concerns regarding the steps in the machine learning deployment framework as described by Ashmore et al. (2021). Retrieved from Paleyes et al. (2020)

.

The model verification stage is partially present at Besi; however, some work on it is still necessary, from internal and external participants. This involvement of the writer is limited to discussion. The first step, requirement encoding with its considerations performance metrics and business driven metrics, is already present. The metrics and KPI's are already there. However, these do not translate to the forecast itself. One metric that is mentioned earlier and measurable is the accuracy, if this is not above a certain threshold then there is no point in utilising the ML model. The regulatory frameworks means that it must adhere to all laws and contracts that Besi has. Seeing as the forecast is communicated to the public, as it is a publicly traded company, there are a set of rules it must adhere to. Although, if the forecast is purely used as a tool for the vice president die attach to apply then the level to which it is applied is the legal consideration. Purely trusting the model and directly communicating that to the public without revision would be unwise. Test-based verification is the testing of whether the model performs as intended. In this case simulations will have to be performed to see whether it can grasp the patterns in the sales data correctly to forecast the future. Furthermore, the data that is used as input must always be validated to make sure that this does not negatively interfere with the result. This consideration overlaps largely with the data management stage and is therefore already covered. Edge case testing is another interesting consideration. It is checking whether the system can operate under extremes. In this case the downturn of both the seasonal and cyclical effect together is the most extreme change to date. It would be essential for the model to be able to spot this change (or even extremer) as people have proven to not be capable of this according to interviews and literature.

Model deployment is a stage in which the focus of this paper resides. Instructions for the integration, monitoring and updating of the system as whole is the core deliverable of this research and thus, will receive extra attention.

Operational support is necessary for the proper integration and utilisation of a ML tool. As explained before, the knowledge about AI models and Besi and organisations in general is limited. Therefore, they will need extra support to be able to apply such a model. This operational support must come from an external AI expert that could explain the ins and outs of the model. In addition, the algorithm should be integrated into a PowerBI system that is easy to use for anyone and automates all the difficult tasks such as training and hyperparameter selection. Reuse of code and models and software engineering anti-patterns are interesting topics for readers with a deeper understanding of AI, but not for business practices. Mixed team dynamics means that a team consisting of the vice presidents of the product groups, representatives of the product groups and AI specialists must form a team. This seems unnecessary. If the model is translated into PowerBI properly then there should be no need for a constant presence of an AI specialist. Proper integration covers more aspects and issues than portrayed in Paleyes et al. (2020). Other sources state acceptation of the new system as a problem Caniato et al. (2011). This is also evident at Besi, where the CEO is known for disliking these models and systems. However, the vice president die attach is enthusiastic and will use the model. Integration will therefore happen, although it may have to go in smaller steps. Where it is used as a simple check first, slowly gaining importance in the forecasting system as it proves its accuracy. End user trust is another issue, but, this is covered in the cross cutting aspects part.

Monitoring and updating are processes that need to be repeated constantly to keep the system working correctly and efficiently. Feedback loops are important addition to the framework as they give insight in to the working of the model. A simple feedback loop such as looking at the predictions the model made and whether these predictions were on average above or below the actual sales can have great benefits. Other feedback loops are further described in 5. Outlier detection is a key tool to flag predictions that cannot be used in the forecasting Klaise et al. (2020). The first way to test this is the aforementioned edge case testing. However, this process has to continue for the whole duration of the use of the model to make sure the model performs as intended. The model is already checked every time it makes a forecast by the vice president so any abnormalities will most likely be detected by this person. Although, it would be wise to write out the ways of detecting abnormalities as the person is not an expert in ML, this is done in 5. Another reason for monitoring is custom design tooling. This means checking whether the model does what the organisation wants it to do and if not to change that. In addition to changing possible unwanted behaviours, looking at how to expand the capabilities of the model is also a strong reason to monitor it. This flows into the concept drift which falls under the updating step. If the model is not behaving as intended anymore, either because the wishes of the company have changed or the structure has changed making the model less effective, then it is time to update it. The model can only predict sales behaviour that mimics the historical behaviour, is it changes drastically then this has to be incorporated into the model.

The cross cutting aspects are an addition by the author of the paper, Paleyes Paleyes et al. (2020). These entail all the aspects that are not directly part of the deployment stages as written by Ashmore et al. (2021), but are important to discuss. Subjects as ethics, law, end users trust and security all come into play in a new context for the organisations implementing a machine learning technique.

When handling data, ethics is almost always a grave concern. The stories of Facebook and Cambridge Analytica are well known and it is unacceptable what these companies do with user data. This is not a concern for forecasting using historical sales data though. No one is directly influenced using private data. There are other ethical concerns though. The model is based on historical data and since it is unknown how the decision making process of a ML technique operates, it being a black box, there could be hidden biases. These hidden biases practically impossible to detect right away; although, constant monitoring of the system enables the user to detect biases in the low run. Accountability, especially when using it as a decision making tool, is an important topic. As stated in 4.1.1, the forecasts are shared with shareholders who base their share price prediction on that. Therefore, the forecast must be accurate and ethical. It would be hard to explain to the public that a black box ML tool predicted a certain number of sales but the actual sales was only half that. This adds to the principle of only viewing the model as a tool and not a decision maker.

Abiding to the law is a priority for publicly traded multinational organisations such as Besi. One wrong move could cost them dearly. Furthermore, being active in multiple countries means having to abide to multiple sets of laws. Ruling out any and all breaches of the laws in all countries for using this tool is best left to experts in the field. However, current laws on machine learning are targeted at protecting personal data and keeping the decision making process transparent Commission (2023). Seeing as the input data is not personalised to people and in fact does not need to be linked to companies either, the first legal concern should not be an issue. The second concern, that of the decision making process, had been discussed. Leaving large decisions to a machine that does not explain its process is a bad idea if transparency is required. Therefore, to adhere to the laws the model should only be seen and used as an additional tool added to the existing system.

End user trust is a big thing as the opinions within Besi about using models are divided, not to mention the opinions of the large shareholders. Getting everyone on board is therefore a task that should be performed diligently. Starting small with the model only being an extra check for the vice president in making the forecast is the best course of action. This way it has minimal impact so there is no reason to object to the use of it which corresponds to the findings in Chakraborti et al. (2017), which say that human managers are more willing to accept new AI technologies if they remain in control of decision making. Furthermore, starting this way means it can slowly prove its effectiveness building trust for the end users. Involving the end users is therefore not necessary from the start, only presenting the value of the tool after running it for some time. User experience is a more difficult task. As explained the users do not poses the skill required to build, maintain and use a complicated ML algorithm. A setup must be made in which the use of the model is made simple and understandable through automatic updates and a business intelligence interface. This business intelligence interface is the way the output of the model is presented to the user and thus an important subject which is further discussed in the development and demonstration phases of this project. The explainability score consideration had been discussed in 2. A black box is impossible to explain, but, limiting its influence and combining it with the judgement of the forecaster makes the final forecast explainable even if the algorithm output is not.

Security is less of a concern to Besi in the first stages of implementing and using the model. The historical data is already largely made public and when using anonymous data to train the model nothing can go wrong. This changes when in a later stage real time sales data from Radar is also used as input for the model. In which case there would be sensitive data in the system meaning the model should be kept inhouse and secure. Besi already works with secure systems to keep its sensitive data secure and forecasts are usually limited to a very select group within the company. If these people are the only ones with insight into the model then the security risk is not increased by use of the model. The other concerns mentioned in 4.2 are of no problem to Besi. Their model is so simplistic that theft is unlikely and if it does happen this will not negatively affect the company. Model inversion is also unlikely as reverse engineering will give no insight into the company practices. Again, the model is a tool so the final forecast will most likely only resemble the outcome of the model not be it exactly. If someone wishes to create a forecasting model based on historical sales data for Besi then they can already do so without needing insider information. This may give a small edge over other investors in the stock of Besi but that edge will not increase due to the implementation of a neural network forecasting tool at Besi.

4.3.3 Conclusion of limitations

Besi, and other organisations, experience many concerns and limitations halting them from implementing machine learning techniques into their business practice. The primary issue being that the field has undergone so much progress in such a short time that most people who have worked their way up the corporate ladder have little to no knowledge about the subject. Further common concerns gathered from a survey of case studies Paleyes et al. (2020), shows many issues that companies have experienced when implementing ML techniques. Most of which concern Besi as well.

The limitations that must be overcome and the concerns that have to be addressed are many. All of these are (partially) solved in chapter 5. Table 4.1 illustrates the fact that the data management stage is already solved. The rest still has to be performed; although, the model learning and model verification stages are to be performed by an external party. The other concerns are all addressed in this paper. Chapter 5 dives deeper into the integration, monitoring and updating of the model and all the cross cutting aspects as detailed in table 4.2. These are all the deployment steps and peripheral matters that that need to be solved for proper and efficient use of the model.

The list of concerns that require extra attention in the following chapter reads as follows:

- Operational support
- Feedback loops
- Outlier detection
- Custom design
- Concept drift
- Aggravation of biases
- Regulations
- Involvement of end users
- User experience
- Explainability

These items are discussed in sections 5.1 and 5.2.

4.4 REQUIREMENTS

When designing a new model or framework one of the first steps is to determine why it is that a new model is needed. What does the organisation want to achieve with the new model and how is that achieved. This section looks at the requirements that Besi has that the new model and framework can achieve and must adhere to. The accuracy of the algorithm, the ease of use, the level of explainability and how easily understandable the output is, etc.

The manager planning and production planner, who have to work with the forecast handed to them by the vice president die attach, stated their desire for an accurate and stable forecast (Appendix .1). The vice president die attach named the high influence of bias due to the human factor as an issue that he hopes will be (partially) solved by the new tool (Appendix .1). The neural network counters the biases with a stable emotionless forecast, as he views it.

The desire is not for it to take over the complete forecasting task. As the vice president die attach sees it, there are too many unpredictable factors that can only be explained through talks with customers (Appendix .1). This is true and should be upheld. The machine learning technique should become a tool that aids the forecasters in their quest for accurate forecasts.

4.5 INPUT AND OUTPUT

For an effective machine learning tool two crucial aspects that have to be dealt with diligently are the input that is going to be used and the output that is expected. For the sake of simplicity of the model having fewer inputs helps. Moreover, in contrast to what some may think, the accuracy of the model could also benefit from less inputs. This is because more data sources could cause the model to overfit on irrelevant aspects. This further researched in section 4.5.1 where in addition to literature, sources from Besi's database are used to find out what data is available for use.

The output is determined by the desired conclusion of the algorithm but it is also important to know how that output is structured. What does the information tell, how is it structured and what other points of interest are there that may be necessary to enhance the results credibility. The desires of the forecaster are most important in this section although literature is also taken into account. Section 4.5.2.

Furthermore, training a neural network is a process that requires a lot of computing power and takes a long time, and that computing power and time is increased drastically when more data is included. Keeping in mind that the objective is to demonstrate and ease the usability of an artificial neural network, it is wise to keep these aspects as low as possible.

4.5.1 Input

The input data must adhere to certain criteria. First, the data must be available and enough of it to train the algorithm. Second, the data must be reliable. Trusting unreliable data is a grave mistake. If the information is in any way skewed, either by inaccurate measurements or with intent then a neural network will spot a very different pattern than the reality. Third, the data must be relevant. There is no point in including data that does not accurately represent the object that is being analysed. This also includes the fact that the data has to be up to date and structured properly.

Besi has been collecting data diligently over many year so the availability is covered. Over these years many things, such as the COVID19 pandemic and the rush for electronics due to remote working, have influenced sales, thereby decreasing the reliability. The data that Besi has collected is still relevant though.

The stored data includes monthly sales, weekly forecasted sales, real time purchase orders, hit rates, etc. All could be included in the model, however, determining the best input or right mix of inputs is the wiser decision. The previous forecasting with neural networks study performed at Besi solely utilised past monthly sales data Steenhuis (2017). The senior vice president global operations BESI and managing director Besi Apac. inquired if it would be effective to not only use historical data but, to also include confirmed future sales into the model as well (Appendix .1).

Utilising the future sales would theoretically benefit the accuracy of the algorithm. However, when training the algorithm, all inputs of the model have to be linked to each other in a universal time step. A time series model with some data described in weekly intervals and others in monthly will not operate properly. Seeing as the most logical input to start with is monthly sales, and this data is only available monthly, all other inputs must be presented as monthly data. This means that the confirmed future sales must be bound to a month. This month, however, is in the future and would overlap with the (output) forecast sales.

Other aspects, shortly touched upon in the beginning of this section, are overfitting and time constraint in training. Data from 2011 is available for certain machines. However, using all that data may result in overfitting of the model in addition to the training of the algorithm taking very long. Enough of the historic data must be included to train the algorithm but not so much as to cause the other issues. Finding this ideal spot is best done via testing. Although, theory could indicate the bounds within which the size of the dataset should fall.

The amount of data that is available in the database of Besi is the upper bound. There is no way of going above that. Literature cannot define a formula which would indicate the ideal dataset size and window for a given problem. However, lower bound can be defined by looking at the seasonal and cyclical patterns. When forecasting a season pattern, the model has to have seen such a pattern in its complete form before Sevey (2017) Brownlee (2019b). If not, then it cannot mimic its behaviour.

In the case for Besi for seasonality this means at least a year worth of data. The cyclical pattern of machine sales at Besi is larger than that though. The cyclical pattern has occured twice in all available historic data. The first ranging from 2014 to mid 2018 and the second from mid 2018 to late 2022. To take both seasonal and cyclical patterns into consideration at least one pattern of each must be included in the training data. Therefore, the lower bound of the training dataset is from mid 2018 to present and the upper bound is from the start of recording till the present.

4.5.2 Output

Besi currently works with forecasted monthly sales. The precision of a monthly forecast is good enough for effective planning. While also broad enough to not force the executor into undesired terrain. Keeping the output data in the same form as the information that is used and communicated in makes it easier for the forecasters to use. Therefore, keeping the output on a monthly forecast seems the right way. The vice president die attach confirmed that this is indeed the preferred way of projecting the forecast (Appendix .1).

So a projection of the expected sales in a certain month is the output. Then what are the peripheral matters that enhance the usability and credibility of this information?

The forecast may give a single number as the expected sales. However, forecasts always come a a probability with a standard deviation. Artificial neural networks are no different. The ML model draws its conclusion based on the most probable outcome but it still has the probability of other outcomes. Including all these probabilities in a manner that is easily understandable and readable would be very valuable.

With that information the forecasters could more easily and clearly interpret the results of the model. If the standard deviation is high then the results are less reliable and if it is low then the models accuracy could be trusted more. Given that the model itself can be trusted and empirical findings do not show great contrasts.

Besi currently presents its forecast data in tabular form, as can be seen in table 2.1. Theory suggests, however, that for sales forecasts it is more effective to utilise graphical displays Harvey & Bolger (1996). This way the forecaster has a better insight in how the sales volume evolves over time. Furthermore, quarterly shifts and delays are clearly visualised in the graph for the next month. In addition, with a graph it is easier to show the uncertainty of the prediction with an upper and lower bound expanded over time. Even better is to make a whole dashboard with insights. According to Choi (2021), AI users need to feel that they have a high level of motivation, ability and clarity. This means that the developers need to create a user interface for the users in which all relevant information is shared with the user. Furthermore, the study shows that for the betterment of trust in the model it is a good idea to include various communication features. For example, graphs, text, tabs, pictures, etc.

Furthermore, there are certain scenarios in which the model will perform better or worse. These scenarios need to be identified before hand to prevent any issues arising from misinterpretation or false forecasts. For example, if scenarios occur that are not represented in the training data. The model is unable to predict outcomes that it has never witnessed before. So, if for example, an unreasonably large order comes in from a single customer then that should be added separately to the final forecast. If the input from Radar is also included in the model, then it should be possible for the model to include these large orders as it can spot the patterns in the actual data, so long as it has historical data which ended in a similar outcome.

It could also be that the model tends to over predict or under predict constantly. This can be resolved in multiple ways. First, the forecaster could use the information of over prediction to counter it by forecasting under it. Another, more clean way, is to include this in the outcome of the model directly so that the forecaster can use it as intended.

Currently the forecast is reviewed and shared by the vice presidents on a monthly basis (Appendix .1). This is often enough according to them as there is not much point in doing it more often. Changes in the near future are almost always due to cancellations by customers which are practically impossible to predict. Plus this task of communicating cancellations is handled by the planners in the fourth tier of 4.2, who address this every week. For forecasts reaching further into the future there is also no point to revising more often. There is no interest in knowing the exact number that will be sold twelve months from now (Appendix .1). In addition to, there being no way of knowing this exact number. Therefore, keeping this forecast update on a monthly interval is absolutely fine.

4.5.3 Conclusion of input & output

To conclude, time series data, spaced out monthly, on historical sales is going to function as the input. In addition to that, during testing of the algorithm it is interesting to also explore using future sales data. Only two cyclical patterns have been recorded and multiple seasonal patterns occur within each of these cyclical cycles. Therefore, testing on two cyclical cycles (upper bound) and one cyclical cycle (lower bound) and comparing results will indicate the ideal input dataset size and window.

For the output, it is best given in a business intelligence environment in expected monthly sales. In this environment there should be the forecast, given in graphical and possibly also tabular if the users prefer that. Also included should be the uncertainty and standard deviation of the forecast so that the forecaster knows how much to trust the model. Further tests with the model and the forecasters need to be done to write up a plan for when the model is unable to forecast due to any reason.

4.6 conclusion

In this chapter, titled Design, five main points were discussed to lay the groundwork for the development of the proposed framework. These topics were: process, technique, limitations, requirements and input/output. All were discussed and researched through interviews and literature research. The results for each section are summarised below.

The forecasting process at Besi has many different layers where the forecast is made more specific in each following layer. The layer of interest for this research is the third layer in figure 4.2. Here, the vice presidents receive the information from sales teams, historical databases and the product groups to make their sales forecast on the total machines for a certain month. That information is passed on back to Radar for official reporting and primary plan for further use within Besi (Appendix .1). There are still many steps Besi has to undertake to effective implement a machine learning forecasting tool into their forecasting process. These steps and their level of completion are depicted in table 4.1. The focus for the development chapter, 5, will fall mostly on the model deployment stage.

The forecasting technique currently applied at Besi can be described as almost purely judgement. To counteract the bias that people tend to have when making forecasts, a machine learning forecasting technique could be very effective. The neural network would utilise historical data and could possibly integrate actual data from the sales teams, Radar, to increase effectiveness. As stated before, a multi-layer perceptron is the ideal ML technique for Besi, and that should be trained using this data. A forecasting support system (FSS) should be set up to aid the forecasters to use the ML technique as a tool in their set of techniques.

Organisations face many limitation in their pursuit of implementing ML techniques into their business practices. Data management is already sorted at Besi. Model learning and model verification are actions to be performed by an expert in the field of machine learning and coding. The model deployment stage and cross cutting aspects still need attention and are the focus in the development chapter. More specifically, the following aspects are to be dealt with: operational support, feedback loops, outlier detection, custom design, concept drift, aggravation of biases, regulations, involvement of end users, user experience and explainability.

The desires of Besi are relatively simple. They wish for a neutral tool to apply in their forecasting process to mitigate the human bias which hampers the accuracy of their forecasts.

The input to the system, as also touched upon in the process and techniques sections, are the historical sales and product group data. Of these, the historical data is the most realistic to apply in a first phase ML technique. Including the actual sales data is an interesting topic to explore but could add extra complications. The output should be represented in the same form as the historical data and the information the forecasters pas onto further layers. This is monthly sales forecast. Furthermore, including information that aids the forecasters in understanding and explaining the results of the ML tool are essential. Things as standard deviation and uncertainty of the results of the ML model are to be included in the design.

The design phase first sketched the current scenario, split into the process and the techniques in place at Besi. Then, the limitations and the requirements with a new system incorporating a machine learning technique were analysed. Lastly, the input and output of the system were described. All of this was done through interviews with relevant players within Besi and accompanying literature reviews. This led to the following conclusions.

A lot still has to be done to realise proper integration of a machine learning technique into the forecasting process of Besi. So much in fact, that further research outside the scope of this paper is required. The topics that are to be addressed in chapter 5 are the process, deployment, cross cutting aspects, input, output and the complete integration of all these facets. Basically, all the parts that Besi has not completed yet that are needed for the usage of a machine learning technique outside of the creation and testing of the model. Furthermore, a business intelligence interface must be developed to increase the ease of use, ease of communication and for the benefit of building trust in the model by the end user.

To complete these tasks collaboration with the end users of the proposed framework is necessary. First, the information gathered in this chapter shall be used to create preliminary designs and solutions. Those designs shall then be presented to the vice president die attach, the senior vice president global operations BESI and managing director and the professors for their feedback. This is an iterative process that will require multiple cycles to eventually reach a valuable conclusion.

5 DEVELOPMENT

Following from the conclusions in the design phase; in this chapter the process (framework), the complementary peripheral matters and the user interface are developed. These were concluded to be the three main focus points in chapter 4. All the parts of this chapter are developed in consultation with Besi and theory acquired in earlier chapters.

5.1 THE PROCESS

The current process at Besi and accompanying literature have been researched in chapter 4. A proposed integration of a ML model into the forecasting process of Besi has been created, visible in figure 4.4. However, this is only a simplified concept of what the whole process would look like and what steps have to be undertaken. Those steps, and more specifically the steps that are new to Besi, are detailed in section 4.3. These two aspects are brought together in a detailed process framework incorporating the machine learning technique in this section.

5.1.1 Integration

The integration aspect deals with the sources, actors, input and output of the system. These have been visualised in figure 4.4 and explained in more detail in section 4.5. Integration covers another essential topic which has not been attended to enough in this paper. Namely, protection of the wider system in case the ML model fails or gives an incorrect prediction.

In the case for Besi there will be no immediate negative impact. The ML model is to be used as an extra tool in the arsenal of the forecaster to base the forecast on. Therefore, if the model fails, the forecaster can go back to the way the process of forecasting that was conducted before the introduction of the model. The current information sources (sales through Radar, historical data knowledge and product groups input) is enough for Besi to make realistic forecasts. The forecasts based on these sources are trustworthy enough to be used as projections for the stockholders of the company. Therefore, Besi has a great fail-safe by falling back into its previous settings as also visualised in figure 4.4.

However, there will be some operational support necessary during the integration phase. This is needed to solve any hick-ups that may occur. The person in charge for the operational support shall be the one making the algorithm as that person is the expert on the algorithm. The tasks this person shall have include: solving any technical issues and explaining the model to the forecaster.

5.1.2 Monitoring

Monitoring of the system will be one of the main tasks of the forecaster once it is integrated. Multiple different aspects will have to be monitored. These include: input, environment, internals and output. There is an argument for including externals in this list as well, however, this is not included in the paper by Ashmore Ashmore et al. (2021). Partly for this reason it is not included here, but it will be touched upon in section 5.1.4, as it falls more under forecasting choices. More specific issues as identified in 4.3 are handled in this section. These topics are: feedback loops, outlier detection and custom design.

The input fed into the system must be monitored as it directly influences the output of the system. Incorrect or unlabeled data could mess up the models results. As it currently stands, the setup for data management of the historical data is good. Everything is labeled correctly and structured properly. Besi Apac is the one updating this and as long as the model does not show any abnormal results then no action is needed. If there are strange results however, then the input data is one of the first things that needs to be checked. Furthermore, the data in Radar is also double checked by the sales teams and the product managers so that does not require constant monitoring either.

The environment that the model operates in is something to take into consideration. For example, the sales during the pandemic, when consumer electronics demand shot up drastically, are not necessarily representative of the real world. Other examples are growth of the economy in general or political issues in the world. These things are impossible for the model to predict. Even for people these things are almost impossible to know. Although, Besi does receive reports about the expected growth of the sector in which they operate. Combining that information with the results of the model and other sources makes the forecast more accurate and relevant, especially for long term forecasts.

Internals are similar to the environment just discussed in that they influence the sales but the model is unable to pick up on it. For example an abnormally large order from a single client would mean many more machines have to be produced, but the ML model cannot forecast this. It may be able to do so if the information from Radar is included. This could skew the data so drastically that the model would not operate properly anymore. This is where the strength of the hybrid forecasting model comes into play. As the judgement by the vice president die attach and other sources of information can counteract or include the abnormalities that the model cannot. Abnormalities are reasonably easy to spot, even for people with a limited understanding of machine learning. If the results of the model differ largely from the information from the other sources, then it needs another look at. This does not necessarily mean that the model is wrong, if it were to always simple state what the other sources say then it would not be useful. It could be that the other sources are over optimistic for example. The general rule should be that if the results differ a lot, first the input data must be checked and if that is all correct then the other sources of information should also be questioned.

Monitoring of the output is simply checking the output to see if anything is wrong. As explained before this is especially useful because there are other sources of information for the forecast with which the model output can be compared.

From the more specific issues regarding monitoring, feedback loops can cause a real head ache. This happens when during retraining of the model, the model is configured in a way that influences its own behaviour based on previous behaviours. If this process is repeated multiple times it could cause the model to develop unwanted habits. The simplest way to prevent this is to train the model separately every single time and only adjust it by hand when necessary. Outlier detection has been talked about above. If any strange results show up then the forecaster must ring the bell and check all the sources. Custom design is not something that has to be done to protect the model or the business from the model. Instead, it is the degree to which Besi or other organisations wish to build upon the model that they have. Through monitoring of what the model does well and what it does not, steps can be taken to make the model even more valuable. In addition, any improvements such as the suggested inclusion of the sales data from Radar can be analysed and included if deemed valuable.

Another thing to look out for is aggravation of biases. This is actually an aspect linked to ethics as applying a seemingly neutral ML program to detect something such as tax evasion has led to minorities being targeted unfairly. The goal is to take human bias out of the equation but, if the data that is used to train the model is biased as well, then the bias only gets amplified. With the case for sales forecasting this is not an ethical concern but the implications are dangerous. Besi wants a neutral model that can challenge the human bias in the system. If the model only enhances those biases then it is worthless. Therefore, this is another aspect that should be monitored diligently; in addition to, the input data needing to be bias free.

5.1.3 Updating

Updating of the system will be necessary to keep it relevant and working, but also to make it better by including more features. To be able to update a machine learning algorithm expertise in the field of AI will be necessary. Therefore, this task does not fall upon the vice president die attach of the product groups. Instead it should be done by an external party, the same party that created the ML model in the first place.

This means that updating could become costly as external experts are required to perform it every single time. Therefore, keeping the frequency of required updates to a minimum is beneficial. To achieve this first of all the model should be working as intended from the beginning. Therefore, a diligent process of development of the algorithm should be undertaken in which it is tested multiple times to prove its effectiveness before proceeding.

Furthermore, the model should be able to retrain itself based on new inputs as time progresses. If this all works properly there should be no problems which need to be fixed by an external party and the vice president die attach can start to come to grips with the model.

Once the forecasters understand and are comfortable with the model, then, it can be updated to increase the accuracy. These updates have been discussed and include adding the information from Radar to the mix of inputs. Other potential updates must be found by constant monitoring of the efficiencies and inefficiencies of the model.

One such inefficiency that could creep up over time, as identified in Paleyes et al. (2020), is concept drift. It is defined as changes in the observed distribution between the input and output of the model. Such changes can occur due to an external even that changes the input data or gradually over time where the environment being observed changes. To prevent concept drift from becoming an issue again monitoring of the system is required. Furthermore, if other sources such as market research show that the market has grown or will grow abnormally fast, then this information can be used by the forecaster to adjust the models output upwards.

5.1.4 Forecasting

The ML model is a tool to be used by the vice president die attach to create a forecast, just as the historical data, sales teams input and product groups input are used. There are multiple sources of information, and with the inclusion of a model, the dynamic of the usage of these various sources changes. This section explains the way the information from the sources interact with each other and create the forecast. In addition, things to look out for when making the forecast in different scenarios and how to respond are detailed.

A short recap; the current forecasting process is one where the information from the sales team is used and discussed with the product groups. A base forecast is formed from this. The forecast can then be scaled up, down or not at all based on historical sales data for that period. The ML model will most probably serve a similar role as the historical data in the beginning as trust has to be built before it gains prominence.

Ideally the forecast made by the model would be trusted fully; however, this is not realistic. The role it can serve from the start is thus as an indication for rounding up or down but also as proof or disproof of the other sources forecasts. More importantly, it can serve as an extra check whether the other data sources are correct or need to be checked again. As the vice president die attach stated, he would appreciate an emotionless model to confirm or challenge the forecast made through the other sources. This is exactly what he is getting.

Thus, the way the ML model should be applied in the beginning stages is to help guide the forecast from the other sources and as a extra confirmation or reason for reassessment of the other sources.

5.1.5 Communicating

After the forecast has been made it has to be communicated to the desired recipients. These recipients are the planners via primary plan and the stakeholders via Radar. They require certain information regarding the forecast and its accuracy and reliability. The ideal ways of communicating the forecast to these two groups is explained in this section.

Currently, the forecast to primary plan is given once a month with a projection of nine to twelve months forward. The data is given as the number of machines that are expected to be sold in each of those months. For the update in Radar it is different. Since that data is used to communicate with the outside world it has to be explainable with a certain degree of certainty. The information sent to Radar is also in a different format than that sent to primary plan.

The framework, designed in section 5.3, is designed in a way that supports clear communication with all relevant parties. Nothing about the communication to parties other than the forecaster should have to change. This way resistance to the new model from those parties will mostly be avoided. So the user interface should be able to easily translate any data to the format that is currently used to communicate with others.

Another point of communication is how the model influence and workings are explained. Some stockholders may be interested to know what the new technology is that is attributing to the forecasts. It is obvious that organisations must always be honest about these things, especially if they are publicly traded. Although, just as with senior management, it does not affect these people directly. Furthermore, no decision making power is shifted as the same forecaster will be the one responsible for making the forecast. Therefore, there is no need to mention it right away. Not mentioning it, or only vaguely by saying the organisation in exploring AI possibilities in its business practices, may be the best course of action. This way there is no extra attention and pressure put on the people working on and with it to perform. The calmness benefits the development of the model and framework until it is ready to be presented to the outside world. It is necessary to always state that the model is used as an extra tool and not given any authoritative powers.

5.1.6 The framework

The framework that flows from the findings is depicted in figure 5.1. It shows almost all the steps that go into the forecasting process at Besi, as explained by the vice president die attach and the senior vice president global operations BESI and managing director Besi APac. (Appendix .1). The layer with the planners is missing as it is not interesting for this case.

Furthermore, the steps that must be completed by each actor in the system are included. These steps are explained in detail earlier on in this section, 5.1. These steps are derived from the input from the interviews and literature. It shows the three information sources that were already present and two added, namely the results of the ML forecast and the external information. All sources are incorporated into a dashboard for easy analysis and forecasting.

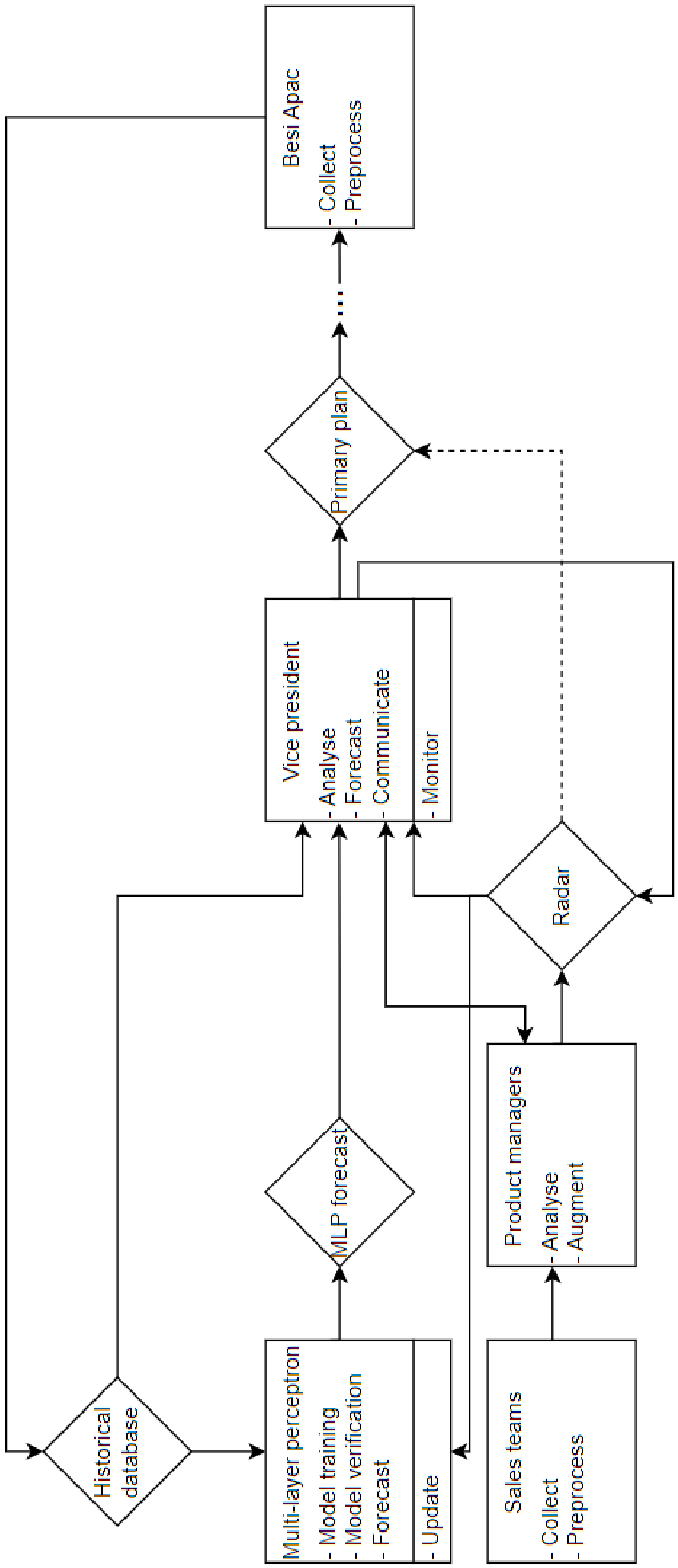


Figure 5.1: The proposed process framework with actionable steps to be performed by the involved actors. Rectangles indicate an actor, ML model included, the diamonds are the information sources and the hexagon is the dashboard.

Figure 5.1 already shows considerable changes compared to figure 4.4 highlighted by the fact that the product groups do not communicate separately with the vice president die attach. Instead the product managers are the head of the product groups and they share all their information on Radar (Appendix .1). Furthermore, seeing as currently extracts from Radar are used to communicate the primary plan to the planners, the idea was generated to enable extracts from the dashboard to be shared directly.

5.2 THE PERIPHERAL MATTERS

Introducing a new way of working to an existing structure not only requires creating the new model. Existing characteristics have to be challenged as well. Subjects such as law and end user trust need to be adjusted to fit the new model. These peripheral or cross cutting matters are discussed and this section.

5.2.1 Law

Discussing the regulations on the subject of AI and ML techniques is difficult as there currently is no clear law on the matter. The EU has been working on their artificial intelligence act for a number of years now, and they are the only ones doing so. Data protection on the other hand has been widely adopted in most places in the world.

Laws on utilising ML techniques for sales forecasting are non existent; furthermore, no such regulation is in the pipeline anywhere. The EU's **EU!** AI act incorporates a number of high risk systems such as critical infrastructure and law enforcement. There is no mention of forecasting in this category. There is therefore, nothing, regulatory wise, holding Besi and other organisations back from applying AI for forecasting. As long as personal data is protected.

Notably the EU is working hard to keep regulations up to date with the technologies. Although they are falling behind, they plan to catch up. Therefore, it is wise to already look at the potential rules that may be applied to the high risk systems. These have no influence on forecasting and will not even be applicable in the first EU AI act, but it could be wise to plan ahead in these things.

The following points are rules which organisations have to abide by in high risk systems: adequate risk assessment and mitigation systems; high quality of the datasets feeding the system to minimise risks and discriminatory outcomes; logging of activity to ensure traceability of results; detailed documentation providing all information necessary on the system and its purpose for authorities to assess its compliance; clear and adequate information to the user; appropriate human oversight measures to minimise risk; high level of robustness, security and accuracy.

Besi complies on almost all of these aspects already or will do so automatically when the model is implemented. The risk is low as the model has no direct power, the mitigation system is the fact that a human is always checking the outcome of the model with other sources. The datasets are of high quality and already used for the purpose of forecasting. Logging of activity can be easily implemented and detailed documentation can be made. The user is given clear information, which shall also be communicated to the others affected. Human oversight is built into the framework. Security is sorted too as the model is kept in house and the data fed into it are also secure on the internal database. The only issues in the model directly are robustness and accuracy. However, any problems caused in either of these two aspects are instantly solved through the interpretation of the forecaster.

5.2.2 End user trust

Establishing user trust is just as important as building the model itself when it comes to machine learning. People not only have to be willing to use the tool, they have to build a degree of trust in the model. This goes for all the affected layers in the company but it starts with the user. Building trust is a process and will not happen instantly; therefore, a plan is directed in this section to help organisations gradually incorporate and accept the model. Topics such as end user involvement, user experience and explainability are all mentioned.

Early on in the report it has been mentioned that the CEO of Besi does not trust models according to his colleagues. Literature research showed that he is not the only one in such a function that has his doubts Yin et al. (2019). Which is not strange, many people in such positions have been working for a long time and are on average older. It can be very hard for these people to understand and incorporate such a disruptive technology. Acceptance of new technologies differs greatly across the different levels and functions within an organisation due to character Gillath et al. (2021) Haesevoets et al. (2021). Although, studies found that as long as the decision making power still belongs to the human agent then there is very little resistance to AI Chakraborti et al. (2017). So long as the AI remains in an advisory role people will accept its introduction. More precisely, Haesevoets et al. (2021) found that so long as a human maintains 70% of the authoritative power, almost all people will accept the introduction of machine generated input. Thus, starting with people who are willing to accept the technology but not giving the technology the power to influence big changes too much from the start is the best course of action.

Yin et al. (2019) also states that there is a big difference between the reported accuracy of an ML model and the perceived accuracy. Both of which are important in building trust in the system. Making a model which has a certain accuracy according to training and testing is different to people applying the model in the real world and experiencing its accuracy. Nevertheless, the first step is to make a model that has a certain degree of accuracy according to tests. People have to feel that it can help them in their decision making otherwise no one will work with it.

Once the model is of a satisfactory level, then it can be introduced into the system in a minor role. This role, as explained before, is as an additional source of information in the the forecasting framework. Not even too much trust is needed in the model as nothing inherently changes. The forecasting process can stay exactly the same and the results of the model can be kept next to it only to see if it works. This way, even if it is not being used, the user can build trust when he or she sees that the model is actually capable of giving an accurate forecast. It is advised to use the model already in this stage though as using the model gives a better understanding of how it operates and and how to use it. Keeping the impact of the model small in the first stages also removes the need to build trust throughout the company in one go, thereby less people have to be convinced of its effectiveness from the start.

So long as the model is only used as yet another source of information on which to base the forecast, no one except the forecaster really needs to trust the model. There are scenarios in which it could come up in conversations though. For example, when the vice president die attach sees that the model gives a completely different forecast than the product groups. In this case the forecaster would have to go back to both sources and validate its information. When that happens the product groups may ask why the vice president die attach does not trust their judgement and the model may come up. However, this will only become a problem if these relays back to the product groups become an nuisance due to the model. If they are not confronted with it every month then their interest and disdain for the model will not hamper any development. Furthermore, The senior vice president global operations BESI and managing director stated everyone is trusted to do their job properly so the forecast that people get handed to them is rarely challenged.

After a while, when the trust has been developed and the model has proven its value, it can be expanded on. Eventually the historical data does not need to be separately incorporated in the forecasting process as it is already incorporated in the model. Once the updates are made where the sales data from Radar is included, the same can apply to Radar. Meaning that in the end only the model and the talks with the product groups are used in the forecasting process. Which is still led by the vice president die attach who has to be cautious of fully trusting the model and monitor its predictions to see if they are believable. This new system would help organise the information and streamline the whole forecasting process greatly. However, it also requires more trust from a wider group of people. In this case not only the vice president die attach needs to believe the model, but also anyone else further down the forecasting chain, the higher ups of the company and external stakeholders. This can only be achieved by the proven worth of the model as these people do not work with the model directly, they only see its results. Therefore, the results must be there before full implementation can be achieved. Thus, creating an iterative cycle of implementation and evaluation, gradually building expanding the model and building trust. The end goal of this is to convince everyone in the organisation, including the CEO who is strongly against such models, that the model is good an trustworthy. Acceptance of senior management is important for the development of AI within organisations Skoumpopoulou et al. (2018) so giving proof of the model working is essential for further investments and developments.

End user involvement is thus implemented from the start as the vice president die attach is the end user. As stated further people are affected by the model and so their experience is of influence too. But, they are not the users so their direct involvement is not required and only the building of trust is important in their cases.

User experience is another aspect that must be stressed. The user must not only trust the model but also be able to operate it and it must feel like the model is making their life easier. On paper it is clearly going to help forecasting and could even streamline the whole process in further stages. Although, in reality, the experience can be very different. That is also why a BI interface is designed in section 5.3. Further thing to be included for the betterment of the user experience are making the model robust to failure. This important for multiple reasons. First, constant breakdowns will frustrate the user and make them resent the model. Second, since the user is a layman when it comes to machine learning, they will not be able to repair the model and thus have to go through the hassle of getting an external party to don it.

Explainability is the last great aspect involved in user experience. It is the ease with which the model can explain itself to the user and the user can explain the results to others. Paleyes et al. (2020) states that explainability is a must have feature to build trust in ML models. This topic was touched upon in section 4.5.2 as the output that the model must be able to generate for the user to understand and use it. Including data such as standard deviation to explain the uncertainty of the model will help the user understand how much to trust the model more. Explaining the thought process of a neural network is impossible but giving the user of the model all the relevant information they could possibly use from the model helps in the explainability. Furthermore, the user must gain a degree of understanding about machine learning such as the fact that it can only replicate what it sees in the data and should not be trusted in every situation.

To conclude, the steps that have to be performed in order to build user trust so that the model is used are the following. First the model has to be developed and proven to work in a simulated environment or on historical data. Then it should be introduced to someone who is willing to accept the new technology but the model should not be given authoritative power. It should be used as another source of information to base the forecast on and to check the other sources. Once the model has proven its worth in business practices as a tool then it can be expanded both in investments in the model and its authoritative power. Furthermore, a basic level of knowledge about machine learning is required from the user.

5.3 DASHBOARD

To present the results of the ML model in an understandable and explainable matter, a business intelligence interface works best. Besi, and other organisations, communicate forecasts by sending tables and graphs to other people within the organisation. Making the machine learning tool able to do that too, in a clear way, is the deliverable of this section.

The desired output, thus what should be shown in the user interface, is that of monthly expected sales and the certainty of the model in its prediction. In addition to that, having all the information in one place could be greatly beneficial to the user. Therefore, including the historical data, sales data and information from the product groups in the interface is explored. Currently at Besi the forecasts are shared in tabular form; although, literature shows that graphical visualisation could aid the user in making more accurate decisions. Various designs according to the literature are developed and demonstrated to the user at Besi to get feedback.

The features that the dashboard could include are:

- Forecast
- Uncertainty
- Previous forecasts
- Historical average
- Sales input
- Market growth
- Largest orders
- Monitoring of previous forecasts

Multiple designs were developed and shared with the vice president die attach. His opinion and feedback is given in chapter 6.3. The examples include a varying array of the features listed above in different sizes and configurations. The examples are simple constructs of different graphs and tables put together in powerpoint. The detail in the example dashboards is limited because it is only a first exploratory step towards a user interface.

5.4 CONCLUSION

This chapter saw the development of the forecasting process framework, detailed descriptions of how to handle the steps in the framework and the peripheral matters, plus the first drafts of a dashboard. All the findings are based on extensive literature research and interviews with people in the field. The developments are shared with the experts in the next chapter, 6, to demonstrate the findings and get feedback for further development.

The process was narrowed down to five main steps; namely: integrating, monitoring, updating, forecasting and communicating. Integration was concluded to be an easy step as the impact of the model is minimal, it does not endanger any other business practices and no extensive training is needed to operate. Monitoring is a continuous process that the forecaster must do to make sure the model is still operating as intended. Updating is split in two different levels. First, updating of the whole model, this is not needed regularly although in the beginning stages it may be wise to try out different configuration to see which works best. Second, the training of the model, this should be done more often, preferably every time the model is used as the latest data can then be applied to it. That would mean a monthly update. Forecasting is performed in the same way it was before only with an extra source of information. The communication does not change drastically either because the other layers in the chain still require the same information from the forecaster.

The framework incurs some drastic changes. It includes a dashboard to present all the information in a structured way in a single place to ease the forecasting job which is still performed by an employee. This is seen as level two automation in the IT domain where the decisions are still made by a person; however, machine learning is applied to ease/automate certain functions Enderle (2021). For a critical function, which also experiences a lot of variation, such as the sales forecasting at Besi it is essential to have a person with experience and extra information make the final call. Thus, the model and dashboard are merely a tool to assist the forecasters in their job.

The peripheral matters that are discussed in this chapter are law, section 5.2.1, and end user trust, section 5.2.2. Laws and regulations surrounding machine learning in the context of this research nonexistent in concrete form. However, looking at the future some aspects may be important to plan ahead in. It is important to stress the fact that the model is not perfect and not in charge a human forecaster will always make the final call. Furthermore, logging of the actions performed by the model and the choices made by the forecaster based on that data could become important in the future for validation purposes and understanding of the model. End user trust is a large topic to be further explored in the demonstration phase. Establishing trust in AI models is easier in certain organisations than in others; Besi seems to be more on the hesitant side of adoption. The building of the trust must be done in small integration steps where the model has little impact in the beginning but could grow out to become more influential as time goes by. First, it must work in theory; second, it is held beside the forecasting process as a confirmation tool; third, it can be expanded further into the organisation.

The dashboard is there to help the forecaster centralise all the information used for making the forecast. There are many sources of information used by the forecaster to make the forecast. However, a lot of these sources are in the back of the mind and thus prone to bias. Having everything central and next to each other makes it easier to make a more precise and unbiased forecast.

All these findings are shared with the vice president die attach and the senior vice president global operations BESI and managing director Besi Apac. in the next chapter. Their experiences in the field help give meaning to the theory by confirming its effectiveness or feedback could be given to improve the current understanding.

6 DEMONSTRATION & EVALUATION

Designing and developing a new framework and accompanying structures based on interviews and theory only goes so far. Demonstrating it therefore, is necessary. The demonstration phase aims to achieve two things. First, the users can get to grips with it to understand how to use it and second, give their opinion on it. Tweaks to the design are logical as business expertise and personal preferences play a role that is hard to simulate through theory and interviews. The framework, peripheral matters and business intelligence interface are all demonstrated to the users in multiple meetings to iteratively improve upon the design.

6.1 FRAMEWORK

The framework, as visualised in figure 5.1, was shared with Besi for a first official demonstration. This first iteration mainly focused on solving any inconsistencies and adding anything that was still missing. The framework was discussed and all the details of each step were lightly introduced.

The first impression from the vice president die attach was good. He stated that all the information was correct and it looked professional. However, the addition of external sources and the dashboard were quickly regarded as unnecessary. The external sources are not directly involved in the monthly forecasting process. Instead they are used for long term planning. Furthermore, the dashboard was seen as an extra to a model that has not proven its worth yet. Therefore, any dashboard implementation would only come after the model forecasting model has been rolled out and been in use for a while.

The result is the framework illustrated in figure 6.1. It shows the first step of implementation where the main goal is to test its accuracy and value as an extra information source. It shows a dotted line going from Radar to the primary plan as the primary plan is an extract of Radar.

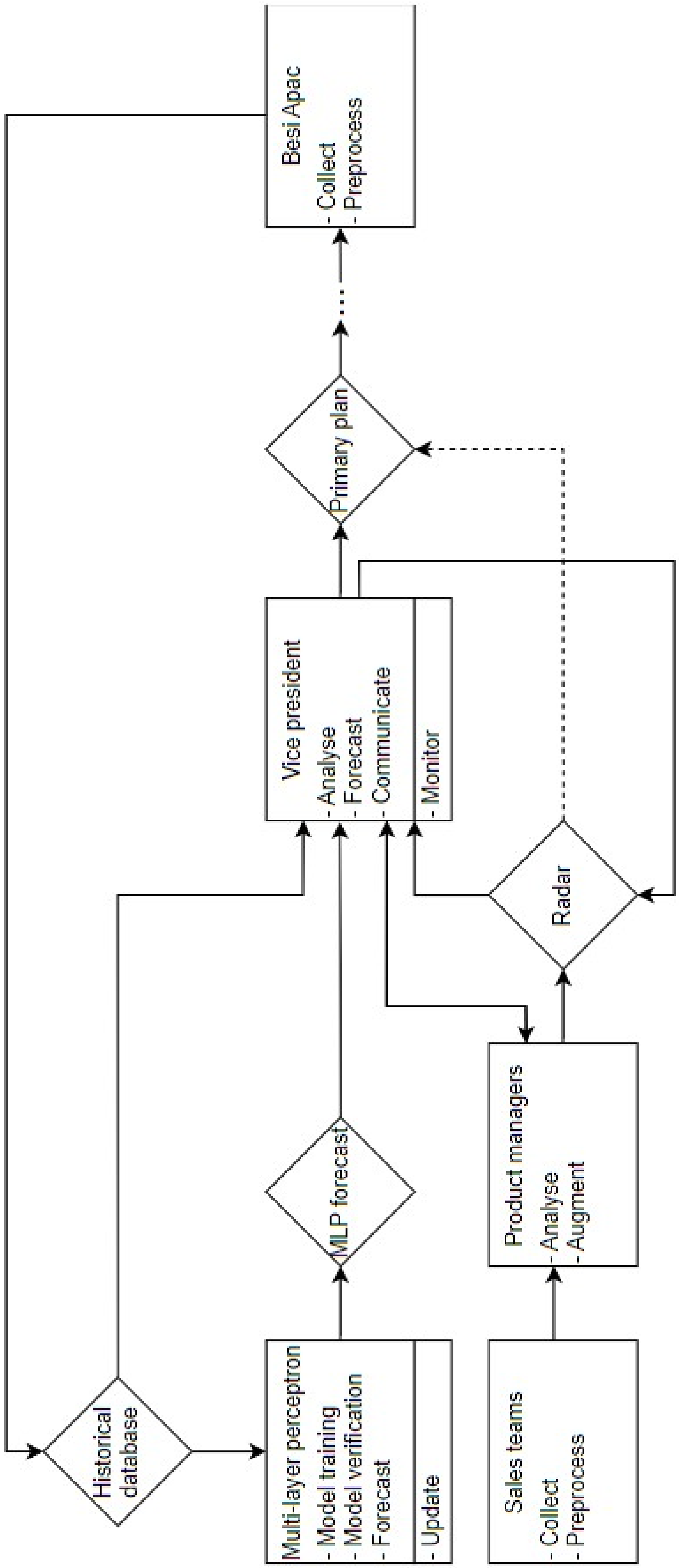


Figure 6.1: The desired (By Besi) process framework with actionable steps to be performed by the involved actors, as a first step implementation of the machine learning technique. Rectangles indicate an actor, (ML model) included, and the diamonds are the information sources.

With this setup almost nothing changes for the forecaster. All the information sources that were present before are still present and represented in the same way as they were. The only difference is the inclusion of a model forecast result. This is the simplest implementation of the model and follows the first step of the implementation plan detailed in chapter 5. Because of both these reasons, it being the preferred option by the vice president die attach and it being the advised first step in the implementation plan, figure 6.1 is the framework choice.

6.2 PERIPHERAL MATTERS

The peripheral matters further explored in chapter 5 are the laws surrounding machine learning implementation and the building of user trust in the model. The other two, ethics and security, that were explored in chapter 4 were deemed unnecessary to develop further for forecasting. The representatives of Besi agreed with this statement. The opinion of Besi on the laws and end user trust, based on the developments in section 5.2, is quizzed in this section. Both the senior vice president global operations BESI and managing director and the vice president die attach gave their opinion on these matters.

Starting off with the subject of law and regulations, The vice president die attach die attach and the senior vice president global operations BESI and managing director feel that the conclusions drawn in the chapter are correct. Although they are no experts on law they do have experience in implementing and developing new forms of doing business so their positive judgement on the matter is a good sign. Also, the inclusion of the 'better safe than sorry' approach is seen as a good idea. With new technologies for which the law is falling behind it is always better to play it safe and for example keep a detailed log of all the workings.

The section about end user trust has a focus on building the trust through the right implementation plan. This implementation plan starts with the framework detailed in figure 6.1 and should, according to the theory lead to the framework in figure 5.1.

Starting with small steps, the minimal viable product principle, means easier implementation, adjustment and lower costs, all of which the vice president die attach sees as positives. Due to the small implementation steps the resistance will also be low allowing for the slow introduction of the new system. Furthermore, the slow implementation has as effect that the trust is built up sustainably and effectively throughout the company. This allows all parties throughout the chain to get used to the system at their own pace and only go to the next step when everyone is onboard.

6.3 DASHBOARD

Literature explains the importance of a dashboard or user interface for proper integration and acceptance of machine learning techniques. However, the demonstration of the dashboards resulted in the conclusion that, at least in the case for Besi it is not desired. In the future it may be interesting to explore this possibility, but for now it is seen as overkill.

The dashboard would be a large investment especially as BESI currently does not use a dashboard. Aside from the development costs in monetary terms the time it takes to design and properly incorporate such a dashboard into the daily practices of the business adds cost. At this stage, with the model not yet having proven its worth, the company does not see a reason to incur these costs. This is fair and is likely the stance of most companies regarding machine learning techniques.

6.4 CASE STUDY

The goal of the case study is to research the behaviour of the user of a ML forecasting tool under correct and incorrect output of the model and/or other sources. The result shows the base level of trust in the model when it is first implemented. It is weighed against the regular sources used for forecasting to see the difference in trust between the old system and the new system which includes the model. The full description of the case study is found in section 3.4.3.

The reason for this experiment is to give forecasters insight into their own bias towards the different information sources. In addition, it adds to the academic knowledge base in the field of trust and bias in machine learning systems versus classical systems. This insight aids the forecaster, and other forecasters, in understanding and assessing their own bias so that they can act accordingly.

The vice president of the die attach branch of BESI, who is responsible for the forecasting activities of that branch, has agreed to cooperate in this test.

The way it is set up is as follows. Historical sales data from 2019, months January through to December, is used as the base line. This year was chosen because it is a relatively stable year with no strange outliers. Using this base year two fabricated sales datasets were created, one 20% above and one 20% under, with a deviation of 2.5% either way. The forecaster was questioned on what he would do in seven scenarios where the normal and two fabricated datasets were mixed up together.

The results were clear in that the forecaster stated in the scenario that both information sources gave the same figure that that figure would be chosen. In the cases where the difference between the two was only 20% then a number roughly in the middle would be chosen. If the difference were 40% then the old data source would be trusted more than the new model; although, still a number between the two would be chosen, only leaning more towards the classical side.

6.5 EVALUATION

Conclusions were drawn about the comments given by the VP die attatch and senior VP global operations. However, this section focuses on the wider implications while the previous section gave first impressions and short reasoning behind those impressions. The results of the case study into bias towards a ML model is also explored in detail in this chapter.

To begin with, the framework had to be altered extensively from the designed and developed framework according to literature to the one preferred by the vice president die attach. The new framework, visible in figure 6.1, is more easily implemented in the forecasting process and thus will meet less resistance.

Literature may indicate that this is not the optimal setup, but, that does not mean it cannot be the ideal first step. When implementing new technologies there will always be resistance, even as in this case from the vice president die attach who is fond of the idea of having a machine learning model aid in forecasting. Minimal investment in terms of cost and time with maximum proof of effectiveness should be the main goal when first implementing a ML model.

Second, the peripheral matters are a broad topic of which only two aspects were explored in depth. There may be other complications in implementing a ML model that were overlooked. However, literature on the most common issues when implementing ML models showed a list of issues and all were either argued as to why they did not need further attention or explored in more detail.

Interestingly even law and regulations did not seem to cause a stir when trying to implement a ML model. This is because the rules on this subject are still being drawn up due to the novelness of the subject. Although, insights into the possible rules and preemptively looking at what the rules could become does give some indications as to how to act. Documenting input and output of the model and how the forecaster used this information is a wise move even though strictly not necessary.

Building user trust is a bigger subject which plays into the final framework, as seen in figure 6.1. It shows, as explained above as well, that people not only need to believe in the value of AI but also experience it before big investments are made. This is in contrast to the findings in section 6.4 where it is shown that the forecaster would trust the model almost as much as the information form other sources. Which implies that the trust in the model, at least for this particular forecaster, is already there, the only thing that is not wanted is the big investment and change in terms of information acquisition.

The dashboard as stated was an unnecessary addition to the research and at this stage does not need further exploring. What the omission of the dashboards means can be evaluated. It shows that though the implementation of a ML model is greeted with open arms, not every single addition to make it better is desired.

The case study showed results that are in line with what is stated above, and as stated, in contrast to the willingness to invest. Three main conclusions are drawn from the results; the forecaster trusts the model as much as the current information sources in low deviation, a slight preference is given to the current information sources when the difference is large and when both sources state the same then that number is chosen.

What this shows is that there is a great level of trust in the model from the start. If the model proves its value in the tests and shows it can accurately predict the sales in a controlled environment then the forecaster will trust it almost as much as the current information sources. This should in theory increase the accuracy Makridakis & Winkler (1983), however, due to the nature of the market in which Besi operates, this may not be the case. The market fluctuates a lot and the model is not able to predict randomness. By taking the average of the two information sources you account for this a little but a more delicate forecast should be made in certain situations. Outliers such as an abnormally large order from a single customer, which may happen, are best forecasted by a human actor as they are more capable of incorporating this information. Meaning that if such a thing were to happen then the forecaster should weigh the classic information source more than the model.

This test was only performed with one test-subject who was already positive about machine learning for forecasting. Therefore, it cannot be assumed that every forecaster will react the same way. If they do then that would mean rapid implementation and adoption of the new technology but it is clear from the fact that this has not happened yet that that is not the case. It is likely that many forecasters would be hesitant to use the tool to a great extent. Even the forecaster questioned for this test stated in earlier interviews that he would only use the tool as a controlling mechanism in early stages of implementation (Appendix .1).

7 CONCLUSION

Literature research and interviews with experts brought forth an ideal design for the implementation of a machine learning algorithm for forecasting in organisations. Talks with forecasters in the field showed that the ideal picture does not exist. Thus a simplified version was constructed and instead a gradual implementation plan had to be devised. This implementation process is explained below in a generalised form.

7.1 RELEVANT INFORMATION

The importance of exploring AI adoption in business cannot be underestimated. Literature research in chapter 2 showed that in many cases a machine learning model can help perform regular business activities to a higher level. However, implementing such a change is a challenging endeavour. All the necessary resources have to be present, the right model had to be chosen, it has to be programmed, implemented, adopted and properly used. Most of these tasks were explored in depth in this research paper.

First, a lot of relevant data is required, plus, it has to be structured properly for the model to be able to use it. Then, the right model should be chosen which is a task that should not be underestimated. This step can be largely done through literature research but also requires trial and error to perfect. Then a skilled developer has to programme the model so that it can be tested and later implemented. The implementation goes through multiple steps in itself as a slow adoption may be required for stiffer companies where resistance to change exists. Proper usage of the model means to know when to trust it more and when to lean away from its forecast under for example random jumps in sales. Not only that, but, tracking the input and output of the model and the influence that had on the final forecast must be stored.

In the case for forecasting, ML techniques can bring great contributions to the accuracy of the forecasts. This is true because in many cases the forecaster will have a bias due to either being in a positive or negative flow. A machine is not susceptible to those emotions. Meaning, that a hybrid model of machine and human would, if applied properly, greatly increase the accuracy of the forecast.

Furthermore, as it stands now there is only one forecaster responsible for a whole series of products, as is probably the case at most organisations. If this forecaster were to fall away then there is no one with the level of expertise required to completely fill those shoes. With a ML model this would be easier as there is already at least an indication of the range in which the forecast should be. Thus, building an extra level of safety into the system through implementation of a ML model. In this case someone with less experience can fall back onto the model more than may be intended so a balance has to be found but at least it will help.

Overall, a machine learning model can bring great prospects to organisation as proven by organisations already applying such models. The proper implementation is something that could use more attention, that is why this paper was written. In section 7.2 the implementation plan as advised in this research is detailed.

7.2 IMPLEMENTATION PROCESS

The implementation process is a step wise process with the final goal shown in 7.3. This section explains the implementation process in terms of the changing forecasting process and the systems(s) in place to support the forecaster. The minimal viable product, the level to which the model should adhere, at each stage is given.

7.2.1 Current

The current system as shown in figure 7.1 is a generalised form of the forecasting system at Besi. This framework has all the basic elements and actors that most forecasting systems at organisations apply. Many organisations currently applying this framework, or a similar version, may wish to implement a new system in which machine learning is included. In that case the implementation is advised to be done in two phases as explained in the next sections.

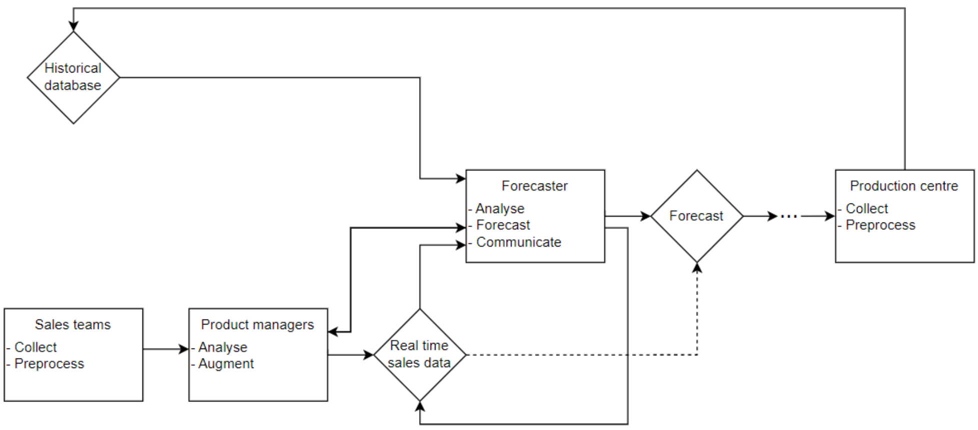


Figure 7.1: The process framework as it currently stands at most organisations; without ML model. Rectangles indicate an actor, ML model included, the diamonds are the information sources and the hexagon is the dashboard.

7.2.2 Phase 1

Phase 1 is the minimal implementation of a machine learning technique. The model has not been granted any power and it is not widespread implemented. In this stage only the results of the model are shown in simple fashion so the the forecaster can use this extra information source as he/she desires.

This is the minimal viable product. A simple model, only utilising the historical data to predict future sales. It has to act with a certain degree of accuracy to be useful; however, fine-tuning of the model is still done when in use. Aside from fine-tuning, this stage is also used to help the forecaster build an understanding of the model and slowly increase the trust. The use of the model is purely as an extra check on the forecast made via the old route. Therefore, there should be no resistance from other levels in the supply chain, although, sharing that there is such a model present helps build trust throughout the other layers at an early stage.

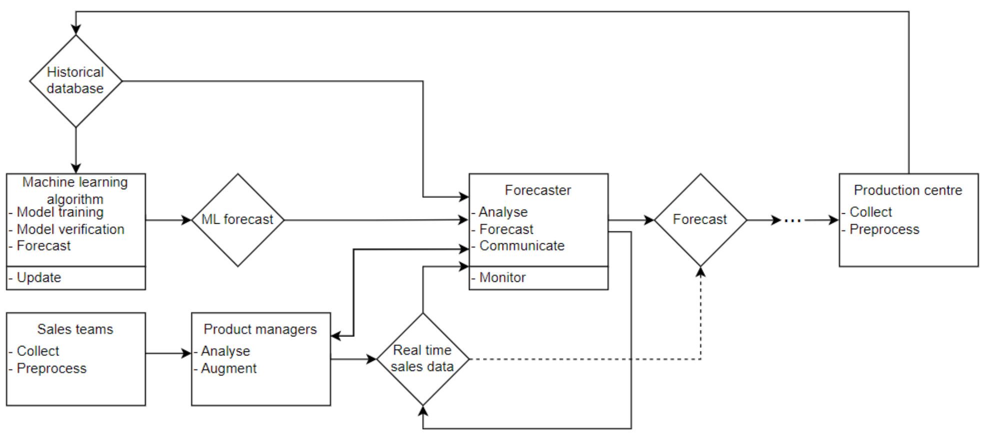


Figure 7.2: The proposed phase 1 process framework with actionable steps to be performed by the involved actors. Rectangles indicate an actor, (ML model) included, the diamonds are the information sources and the hexagon is the dashboard.

There is thus only a slight change in the forecasting process; but, the model has to be created from scratch. For this to be possible multiple actions have to be undertaken and a certain level of infrastructure and data must be present. Paleyes' paper Paleyes et al. (2020) details these challenges well, an overview of these challenges is visible in table 4.2. Once that is all in place someone with knowledge on creating machine learning algorithms can create a simple version to start working with.

Such a model would have to communicate the forecast in a way the forecaster can understand. Therefore, preferably in the same format as the forecaster shares the forecast to other levels in the chain. The ideal algorithm to use varies per case in terms of difficulty to develop and accuracy. In the case of Besi a MLP was determined to be the best solution, chosen for its simplicity and effective accuracy in forecasting in similar cases.

7.2.3 Phase 2

In the first phase the focus is on making a working model and proving to the forecaster that it is an effective tool. The next phase is to roll it out and expand it into the theoretical best implementation form, which is shown in figure 7.3. It features a dashboard to ideally convey the results of the ML model in addition to, utilising real time sales data to make its forecast. The real time sales data enables the model to better flow with changing market trends however it also reacquires the retraining of the model every month.

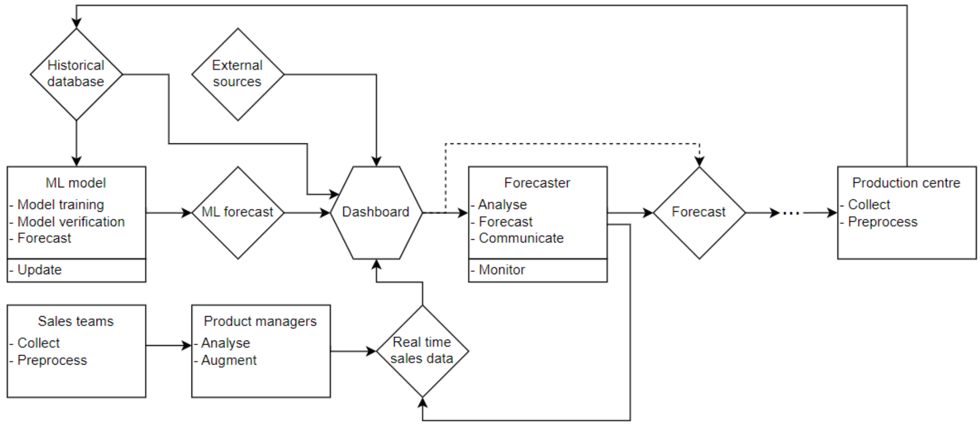


Figure 7.3: The proposed phase 2 process framework with actionable steps to be performed by the involved actors. Rectangles indicate an actor, (ML model) included, the diamonds are the information sources and the hexagon is the dashboard.

The goal in phase 2 is to create a complete overview of all the relevant data for the forecaster to use to make his/her forecast. The complete dashboard is the closest to full automation without giving the model authoritative power. In some cases it may be more efficient and effective to have the model automatically make the forecast and share that down the chain. However, in many cases, such as the one researched at Besi, it is wise to have a hybrid system in which there is still a skilled person making the final decisions.

7.3 LINK TO LITERATURE

This research conducted at a multi-national organisation, BE Semiconductor Industries, gave interesting practical insights into with which to test and add to the existing literature. The consensus from literature was that a ML model could enhance the accuracy in forecasting, however, proper implementation through a hybrid model was less researched. Caniato, Caniato et al. (2011), did dive deeper into this aspect, as seen in section 2.2.7 explaining that proper integration is necessary for acceptance.

The implementation technique the Caniato uses in his work is the 'unfreeze, change, freeze' developed by Lewin Lewin et al. (1947). In Caniato's paper this technique worked effectively in implementing the quantitative tool into a qualitative forecasting system. However, it does not exclude other implementation techniques from working either, such as the one suggested in this paper.

In contrast to the 'unfreeze, change, freeze' technique, the 2 phase technique has a less drastic impact on the current system. Caniato explains the importance to proper implementation to reduce resistance and allow everyone in the organisation to get comfortable with the new forecasting method. Therefore, including an implementation step where there is no direct impact to the current system, instead solely an additional tool, seems a logical advancement on the current theory. Results in this paper, although not thoroughly tested, indicate that such a two phase implementation technique is prefered by forecasters.

This has two main implications for literature. First, it shows scholars that there are multiple ways the implementation of a qualitative method, including machine learning ML, are possible. Although, both are grounded in the idea that a hybrid system is prefered. Second, this research adds another implementation method to this mix. A methods wherein the system is left as it is while introducing a pilot of the new method to test and prove its value before fully integrating a finalised model into the system.

8 REFLECTION

In this chapter three topics are discussed; these are: the scientific and societal contributions of the paper and a critical reflection of the research performed. This reflection is important as the goals of the thesis were to have a meaningful contribution on scientific and societal scales. In addition, learning to perform research and performing a project of such scale is another objective of the thesis. Thus, a critical reflection on what the writer believes went well and what should have gone better must be included.

8.1 SCIENTIFIC CONTRIBUTION

For a master thesis Complex Systems Engineering & Management it is required to make a contribution to academics with the research. This section will go on to explain how this paper has achieved that.

A great number of research papers address the subject of machine learning models applied in various situations. So also for forecasting in different markets and at different companies. What seemed to be missing in many of these papers is insight into the implementation of these models into the complete forecasting process. That is where this research finds its knowledge gap.

The papers invested in the topic of ML models were almost all focused on finding the most accurate model for a certain task. Many of these papers were explored in chapter 2 to find the theoretically ideal model for this case. These papers are important because many different cases have to be explored to provide a good knowledge base. The issue with those papers is that they do not go further in explaining how the implemented their model and how it should be used in a hybrid fashion.

Therefore, this paper creates the bridge between all those papers with the organisations. The step for organisations to implementing a ML model can be daunting. Even at Besi a study into the best model for forecasting was performed in 2017 and it had not been touched in five years time, until this research. Showing how difficult it is to bring such technologies to the forefront and actually implementing and adopting them.

This paper bundles subjects as common issues in implementing a model and the solutions to those problems, implementation steps acknowledging the need to build trust, how to use the model, what data should be used, peripheral matters that could have influence on the adoption of a ML model and a study into the trust a forecaster would have in the model.

8.2 SOCIETAL CONTRIBUTION

In addition to the scientific contribution, a societal contribution must be achieved through this research. In this case the organisation, Besi, is the main beneficiary in societal regards. However, the findings also have greater societal implications as the framework is not limited to only being viable to one organisation. The generalised implementation plan could benefit many more organisations exploring the opportunity to implement machine learning into the forecasting process.

After not finding a good knowledge gap for the initial project at Besi, namely that of optimising the transport network, a new subject had to be found. Through interviews with employees at all levels of the company, CTO, vice president, planning, acquisition, etc. a new topic was found. Many of the issues could be traced back to inaccurate forecasts. Therefore, a solution for this problem had to be found, and one that is proven through literature is that of machine learning algorithms. In 2017 someone at Besi had had the same idea and proved that a MLP model would indeed help increase the accuracy of forecasts. The problem here was that it had not been implemented in the next five years. Thus, the contribution of this paper in societal matters is that it gives Besi another stepping stone in their path to adoption of a ML model, namely the implementation.

8.3 CRITICAL REFLECTION ON PERFORMED RESEARCH

This section will reflect on the way the writer performed the work that went into this paper. The methods and structure of the research are challenged to see if a better option was possible given the luxury of hindsight. Any mistakes made in the performing of the research and a small part on the lessons learned from performing the research are all described in this section.

8.3.1 Problem statement and objective

The problem statement, section 2.3, and objective, section 3.1 were developed at the Besi Apac. office after the initial research proposal was denied. The problem statement was developed through interviews with numerous employees of Besi Apac. of different sectors and layers of the company. Each person had their own thoughts about what was wrong and could be better, all these thoughts were organised and a theme that was seen throughout the business was identified. This theme was the of incorrect and often changing sales forecasts.

The next step was to find the right way to resolve the issue. Therefore, the chain of forecasting was detailed and visualised in figure 4.2. This visualisation together with interviews allowed for the identification of the location of maximum impact through a CoSEM thesis research. This was almost at the base of the forecasting chain as all the forecasts after that are based on this one.

Prior research at Besi APac. had shown that an artificial intelligence model could be utilised to improve the accuracy of forecasts. Thus, the next logical step is to write the implementation plan. This was the objective of the research. In hindsight this was the right course of action based on the talks with employees and literature research.

The identification of the problem and the pursuing objective were done in a diligent fashion in constant consultation of the managing direction of Besi Apac. and the supervising professors of Delft university of technology.

8.3.2 Research approach

The research approach, meticulously detailed in chapter 3, was made to structure the research in the best way possible to reach the research objective. The way the research should be performed was also researched in detail. Eventually the design science research (DSR) approach was chosen. The reason for this was that 'there is a gap in the knowledge of the functioning of a (complex) socio-technical system. Furthermore, a design in the form of a framework has to be made and proven to work.' as stated in section 3.2. This held true throughout the research, thus, the reasoning behind the choice of approach was correct. However, whether it was the best approach to go by is another question.

For a master thesis of Complex Systems Engineering & Management there a typically six different approaches possible, according to the TU Delft course 'Master thesis prep'. These approaches are: case study, quantitative research, modelling, exploratory, mixed methods and the design research approach. When looking at all the options again it is clear that the DSR approach was the right choice for this research for the simple fact that the goal was to design something. All other approaches aim to answer a research question and develop an understanding of a new or existing phenomena. The DSR approach takes existing information to plug a gap in the knowledge of a socio technical domain by designing a solution.

Therefore, following the structure of the design research approach, the chapters were set up in the steps highlighted in figure 3.1. These chapters are titled; Design 4, Development 5 (these were split for clarity reasons), Demonstration & Evaluation 6 (these were added together for clarity) and Conclusion 7. This was a good structure to work with. Every chapter flowed nicely into the next providing a foundation on which the next chapter could be built. The steps were performed properly and done well, with feedback and iterations throughout all steps instead of only from 'evaluation' and 'communication' back to 'define objectives of solution' and 'design & development'. Instead, most of the feedback was incorporated back into the development after the demonstration step.

8.3.3 Research steps

The research steps as explained in section 3.4, changed multiple times throughout the research process. This was due to new findings that were seen as interesting additions to the research and because of changing wishes from

Besi. The steps from the design science research approach were broadly followed and only slightly adjusted to fit this specific research. However, new steps and topics were introduced as research indicated its importance to the subject.

The design step was carried out as planned. The steps: Current forecasting process, Current forecasting techniques, Limitations, Desires and Input and Output were all used to build a base of knowledge. The goal was to gather all the information that Besi already had and combine it with literature to find the next ideal course of action. The chapter ended up being a great source for the development chapter to be built on.

The development stage, which was split up from the design stage, bore the objective that was set for the chapter, namely that of a working framework for incorporation of a forecasting tool. It also explained the important steps, as delineated in the design chapter, in greater detail. Thus, the methods applied were good. However, the chapter is a little hectic with many different aspects being tackled. In the future this could be better built up and thought out or parts could even be left out of the scope. For example the dashboard, theory explained it would be valuable but Besi did not desire such a thing being developed.

The demonstration & evaluation chapter was a step where the research, that had always been done in consultation with the VP die attatch, was officially presented and discussed. The thoughts of the VP die attatch and senior VP global operations were collected and necessitated further design and development. The evaluation step warrants process iteration; so, in this research the choice was made to include a small evaluation in the demonstration step and use that to rework the design. This was done as every step along the way had had iterations going back to previous chapters and this way the communication with Besi was clearer and easier.

The goal of the conclusion is to wrap up the whole research and draw a conclusion. In this case the conclusion is almost more of a summary as all the information was already present from earlier in the paper. The change is that in the conclusion it is generalised for a wider audience while in the earlier chapters is was all about Besi. The chapter succeeded in reaching the desired research objective set out in chapter 3.

8.4 relevance to complex systems engineering & management

Relevance to the MSc Complex Systems Engineering & Management is demanded by the Delft University of Technology for this thesis. Thus, this section explains the relevance to the subjects followed and the methods learned during the master. In addition, the relevance to the Transport & Logistics track is explained.

The reason it passes as a CoSEM subject is because it is a truly sociotechnical and multidisciplinary problem. The requirement of artificial intelligence and machine learning knowledge supplies the technical side. Meanwhile, the implementation of the model into the business practice through working together with employees is the social aspect. Thorough analysis of the current situation and the effects of introducing a new technical component was necessary. The framework, in figure 6.1, embodies the essence of socio-technicality with the intertwining of machine and person, in an interdependent, system to reach a valuable conclusion.

The subjects that were most influential in the completion of this thesis were: Master thesis prep, Complex systems engineering, Introduction to design in complex systems and Digital business process management. These courses formed the base of knowledge necessary to carry out this research in all its complexity.

Furthermore, it suffices as a transport & logistics subject as forecasting is one of the main aspects of logistics. Without good forecasting either shortages occur or there is too much inventory which also costs the business greatly. There was no specific subject in the master that taught the contents of this thesis; however, the courses: Emerging and breakthrough technologies and Digital business process management did introduce the subject of machine learning for forecasting.

8.5 LIMITATIONS

This section explains the limitations experienced by the writer during the research.

First of all, the model was never created. This research only focused on the implementation of a model, based on expert opinion and theory, which, in theory, would benefit the forecasting process. However, without proof from the actual implementation of such a model both the implementation plan and the expected accuracy increase are not proven concepts. This limits the conclusiveness of the findings.

8.6 FUTURE RESEARCH

This research was a sequel to a previous research performed at Besi, Steenhuis (2017). That research showed that a MLP would be able to improve upon the current forecasting accuracy at Besi. This research worked with that knowledge to find the ideal implementation plan for the model. Future research should focus on building a working model and incorporating that into the forecasting process at Besi. This is needed fur multiple reasons, first, Besi wishes to continue on this path and actually start working with a forecasting model, second, to prove the theory stating that it would work and third, to prove that the implementation plan works or to improve upon it.

BIBLIOGRAPHY

Acar, Y., & Gardner Jr, E. S. (2012). Forecasting method selection in a global supply chain. *International Journal of Forecasting*, *28*(4), 842–848.

Anderson, D. L., Britt, F. F., & Favre, D. J. (2007). The 7 principles of supply chain management. *Supply Chain Management Review*, *11*(3), 41–46.

Armstrong, J. S. (1986). The ombudsman: research on forecasting: A quarter-century review, 1960–1984. *Interfaces*, *16*(1), 89–109.

Armstrong, J. S., & Collopy, F. (1998). Integration of statistical methods and judgment for time series forecasting: Principles from empirical research. *Forecasting with Judgment, Chichester, England: Wiley&Sons*.

Arvan, M., Fahimnia, B., Reisi, M., & Siemsen, E. (2019). Integrating human judgement into quantitative forecasting methods: A review. *Omega*, *86*, 237–252.

Ashmore, R., Calinescu, R., & Paterson, C. (2021). Assuring the machine learning lifecycle: Desiderata, methods, and challenges. *ACM Computing Surveys (CSUR)*, *54*(5), 1–39.

Besi, B. S. I. N. (n.d.). *Besi company profile.* Retrieved from https://www.besi.com/company/company-profile/ (accessed: 19/10/2022)

Brownlee, J. (2019a). Comparing classical and machine learning algorithms for time series forecasting. *Machine Learning Mastery, Australia*.

Brownlee, J. (2019b). *How much training data is required for machine learning?* Retrieved from https://machinelearningmastery.com/much-training-data-required-machine-learning/ (accessed: 23/11/2022)

Caniato, F., Kalchschmidt, M., & Ronchi, S. (2011). Integrating quantitative and qualitative forecasting approaches: organizational learning in an action research case. *Journal of the Operational Research Society*, *62*(3), 413–424.

Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, *184*(3), 1140–1154.

Carbonneau, R., Vahidov, R., & Laframboise, K. (2007). Machine learning-based demand forecasting in supply chains. *International Journal of Intelligent Information Technologies (IJIIT)*, *3*(4), 40–57.

Chakraborti, T., Kambhampati, S., Scheutz, M., & Zhang, Y. (2017). Ai challenges in human-robot cognitive teaming. *arXiv. Artificial Intelligence.*.

Chambers, J. C., Mullick, S. K., & Smith, D. D. (1971). *How to choose the right forecasting technique*. Harvard University, Graduate School of Business Administration Cambridge, MA.

Chase, C. W. (2016). *Next generation demand management: People, process, analytics, and technology*. John Wiley & Sons.

Chatfield, C. (2000). *Time-series forecasting*. Chapman and Hall/CRC.

Chaudhuri, K. D., & Alkan, B. (2022). A hybrid extreme learning machine model with harris hawks optimisation algorithm: an optimised model for product demand forecasting applications. *Applied Intelligence*, 1–17.

Choi, Y. (2021). A study of employee acceptance of artificial intelligence technology. *European Journal of Management and Business Economics*.

Commission, E. (2023). *Regulatory framework proposal on artificial intelligence.* Retrieved from https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai (accessed: 17.01.2023)

Enderle, R. (2021). *Ai and the 6 levels of it automation.* Retrieved from https://www.datamation.com/artificial-intelligence/ai-6-levels-it-automation/ (accessed: 11.04.2023)

Eroglu, C., & Croxton, K. L. (2010). Biases in judgmental adjustments of statistical forecasts: The role of individual differences. *International Journal of Forecasting*, *26*(1), 116–133.

Feizabadi, J. (2022). Machine learning demand forecasting and supply chain performance. *International Journal of Logistics Research and Applications*, *25*(2), 119–142.

Gillath, O., Ai, T., Branicky, M. S., Keshmiri, S., Davison, R. B., & Spaulding, R. (2021). Attachment and trust in artificial intelligence. *Computers in Human Behavior*, *115*, 106607.

Haesevoets, T., De Cremer, D., Dierckx, K., & Van Hiel, A. (2021). Human-machine collaboration in managerial decision making. *Computers in Human Behavior*, *119*, 106730.

Harvey, N., & Bolger, F. (1996). Graphs versus tables: Effects of data presentation format on judgemental forecasting. *International Journal of Forecasting*, *12*(1), 119–137.

Hofmann, E., & Rutschmann, E. (2018). Big data analytics and demand forecasting in supply chains: a conceptual analysis. *The International Journal of Logistics Management*.

IBM, C. E. (n.d.). *Machine learning.* Retrieved from https://www.ibm.com/my-en/cloud/learn/machine-learning (accessed:20/10/2022)

Khan, M. A., Saqib, S., Alyas, T., Rehman, A. U., Saeed, Y., Zeb, A., ... Mohamed, E. M. (2020). Effective demand forecasting model using business intelligence empowered with machine learning. *IEEE Access*, *8*, 116013–116023.

Kit, A. (2022). *How amazon uses machine learning?* Retrieved from https://www.geeksforgeeks.org/how-amazon-uses-machine-learning/

Klaise, J., Van Looveren, A., Cox, C., Vacanti, G., & Coca, A. (2020). Monitoring and explainability of models in production. *arXiv preprint arXiv:2007.06299*.

Kochak, A., & Sharma, S. (2015). Demand forecasting using neural network for supply chain management. *International journal of mechanical engineering and robotics research*, *4*(1), 96–104.

Lewin, K., et al. (1947). Group decision and social change. *Readings in social psychology*, *3*(1), 197–211.

Lorente-Leyva, L. L., Alemany, M., Peluffo-Ordóñez, D. H., & Herrera-Granda, I. D. (2020). A comparison of machine learning and classical demand forecasting methods: a case study of ecuadorian textile industry. In *International conference on machine learning, optimization, and data science* (pp. 131–142).

Ma, S., Fildes, R., & Huang, T. (2016). Demand forecasting with high dimensional data: The case of sku retail sales forecasting with intra-and inter-category promotional information. *European Journal of Operational Research*, *249*(1), 245–257.

Makridakis, S., Hyndman, R. J., & Petropoulos, F. (2020). Forecasting in social settings: The state of the art. *International Journal of Forecasting*, *36*(1), 15–28.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PloS one*, *13*(3), e0194889.

Makridakis, S., & Winkler, R. L. (1983). Averages of forecasts: Some empirical results. *Management science*, *29*(9), 987–996.

Moroff, N. U., Kurt, E., & Kamphues, J. (2021). Machine learning and statistics: A study for assessing innovative demand forecasting models. *Procedia Computer Science*, *180*, 40–49.

Murphy, K. P. (2012). Machine learning: a probabilistic perspective mit press. *Cambridge, MA*.

Paleyes, A., Urma, R.-G., & Lawrence, N. D. (2020). Challenges in deploying machine learning: a survey of case studies. *ACM Computing Surveys (CSUR)*.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, *24*(3), 45–77.

Punia, S., & Shankar, S. (2022). Predictive analytics for demand forecasting: A deep learning-based decision support system. *Knowledge-Based Systems*, *258*, 109956.

Reifman, F. E., J. (2020). Computers operations research. *Elsevier*, *29*, 1237-1250.

Reifman, J., & Feldman, E. E. (2002). Multilayer perceptron for nonlinear programming. *Computers & Operations Research*, *29*(9), 1237–1250.

Rožanec, J. M., & Mladenić, D. (2021). Semantic xai for contextualized demand forecasting explanations. *arXiv preprint arXiv:2104.00452*.

Severtson, R. B., Franks, L., & Ericson, G. (2017). What is the team data science process? *Microsoft Azure*.

Sevey, R. (2017). *How much data is needed to train a (good) model?* Retrieved from https://www.datarobot.com/blog/how-much-data-is-needed-to-train-a-good-model/#:~:text=the%20time%20series.-,For%20most%20time%20series%20applications%2C%20this%20means%20that%20the%20submitted,of%20the%20maximum%20expected%20seasonality.&text=For%20example%2C%20if%20you%20have,to%20train%20a%20successful%20model. (accessed: 23/11/2022)

Sharot, T. (2011). The optimism bias. *Current biology*, *21*(23), R941–R945.

Shearer, C. (2000). The crisp-dm model: the new blueprint for data mining. *Journal of data warehousing*, *5*(4), 13–22.

Skoumpopoulou, D., Wong, A., Ng, P., & Lo, M. (2018). Factors that affect the acceptance of new technologies in the workplace: a cross case analysis between two universities. *International Journal of Education and Development using ICT*, *14*(3).

Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting*, *36*(1), 75–85.

Steenhuis, T. (2017). Demand forecasting using machine learning techniques in the semi-conductor industry.

Yin, M., Wortman Vaughan, J., & Wallach, H. (2019). Understanding the effect of accuracy on trust in machine learning models. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1–12).

.1 INTERVIEW VALIDATION

The following is a validation of the statements made by the interviewed persons in this paper. Not all talks were recorded and documented. Especially the small short talks where only a few questions were asked when walking by the office gave a lot of insight used in this research. As for those this statement is a validation that all of the comments with a reference to this appendix were said by the mentioned persons.

The following people were mentioned in this regard:

Senior VP global operations Besi & managing director Besi APac.
Vice president die attatch
Manager planning
Production planner

TU Delft