FORECASTING PROMOTIONAL DEMAND VOLUME

by

J.S. (Jelle) Kerkdijk

Master Thesis

Examination Committee

Supervisor Unilever:
Drs. P.D.J. van Balkom

Supervisors UT:
Dr. E. Topan
Dr. I. Seyran - Topan

26th of September 2019

Department of Industrial Engineering and Business Information Systems
Master of Science in Industrial Engineering and Management
Track: Production and Logistics Management

UNIVERSITY OF TWENTE.

All models are wrong, but some are useful.

George Box, 1976

Preface

This master thesis is the result of the final part of my master study Industrial Engineering and Management at the University of Twente in Enschede. The master thesis was executed at Unilever Netherlands in Rotterdam from the beginning of April 2019 to the end of September 2019.

Finishing my master thesis marks the end of my student life and is the start of a new phase in my life in which I hope to keep challenging myself and to explore great opportunities.

I would like to grab the opportunity to express my gratitude towards the people who have helped me during my master thesis and my student life in general.

First, I would like the thank my first supervisor Engin Topan for his support during my master thesis. I am thankful for the time you took for me to give me your honest feedback and for the conversations we had on both academic and personal level. Besides, I would like to thank my second supervisor Ipek Seyran-Topan for your useful feedback, giving your perspective and providing me with the structure I needed in both the thesis and in our discussions.

Second, I would like to thank my company supervisor Patrick van Balkom for the insights he provided me during our discussions. His wisdom provided very useful insights and his guidance shed light on my path the moments I needed it. I am really gratful for the opportunity he gave me to execute my master thesis project at Unilever. I enjoyed working with him and experiencing the business perspective compared to the scientific perspective which was the main focus during my studies.

Third, I would like to thank my father, Henk, for all the time and effort he spend helping me and providing me with feedback.

Lastly, I would like to thank my family, close friends and my girlfriend for helping and supporting me throughout this project and during my entire student life.

Jelle Kerkdijk
Rotterdam, September 2019

Management summary

This research for the master thesis is performed at Unilever Netherlands in Rotterdam. Unilever is a global company selling Fast Moving Consumer Goods (FMCGs). Characteristics of FMCG market are competitive, high volumes, fast turn over, multiple product innovations. In the FMCG market the effect of promotions on sales volume is substantial.

Achieving the highest product availability at the lowest costs in combination with high volumes and high inventory turnover is a big challenge. Because of this, Unilever strives to deliver the right amount of products at the right time to the right customer at the lowest cost, which is only possible with an accurate forecast of the demand volume.

This emphasises the importance of an accurate forecast. That is why Unilever wants to shift towards automating demand volume forecasts of promotions using predictive modelling. Unilever has co-developed a predictive modelling tool, based on a machine learning algorithm.

In the implementation phase the company encountered resistance from the employees responsible for the demand forecast, because the current predictive model does not perform well at their perspective at detailed item level. As consequence, these employees modify data, resulting in an incorrect use of the workflow, leading to lower performance on forecast accuracy in the operation. Low accuracy leads to high stock costs or low fill rate do to underforecasting or overforecasting. This results in the volume forecast for product promotions not being accepted by those employees and therefore not used, which is the main problem in this research.

This research is part of continious improvement programm within the company. The goal of this research is to answer the main research question: “How to improve the current forecasting method for product promotions to ensure it is accepted and internalized by the users?”

Initial analysis of the root cause indicated three main problem areas: data availability, data quality, and the forecast model. The largest improvement potential using the Analytical Hierarchy Process (Saaty, 1980) improving data availability. Besides, the focus is also on the acceptance of the forecast. Therefore, the aim is to get the highest outcome on forecast accuracy in business operation, using the equation in Figure 1.

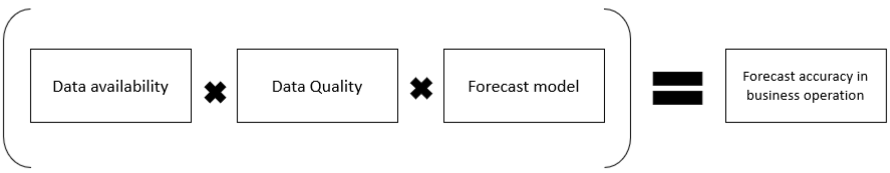


Figure 1: Equation to measure the impact of forecast accuracy in business operation.

The users of forecasting model are interviewed: they feel information overloaded and they find the current forecasting model too complex. According to scientific research, amongst others by Van Loo (2006) and Makridakis, Spiliotis, & Assimakopoulus (2018), indicated that simpler models might be sufficient.

As result of this study:

- A simplified forecasting model for product promotions is built and validated.
  It meets the requirement of at least equal performance to the user forecasts. This model uses less parameters and simpler relation formula's which increases the understandability by users.

- An improved user interface is built.
  With this only the most important variables are used which reduced the information overload of users.

- A monitoring dashboard
  With this dashboard the data availability is monitered. This dashboard is already frequently used by the company.

The implementation of the simplified forecasting model has proven to have a substantial improvement in forecast accuracy in business practice for Unilever can be obtained by improving the data availability of promotions.

Therefore the following is recommended to Unilever:

- Implement the three improvements; simplified model, improved user interface and monitoring dashboard.

- Perform periodic evaluation with two perspectives:
    - Use and acceptance by users
    - Effectiveness of the model

Based on the outcome new adaptions can be done. This report with fundamental research can be used as reference for selecting appropriate measures. It is not a final phase because it doesn't ensure acceptance and internalization by the users yet, but it is a first step in continuous improvement process. It gives better preconditions because it 'overcomes' the major drawbacks of the current forecasting method.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

This chapter gives an introduction to this master thesis. First, the research position within the company is described. Second, a general introduction of promotional demand forecasting at Unilever is given. Third, the structure of this research is described. Fourth, the research is motivated. Fifth, the research questions are formulated. Finally, the scope and practical requirements of the research are stated.

1.1 Company introduction

Unilever is a global company producing and selling fast moving consumer goods among others foods, beverages, cleaning agents and personal care products. Unilever makes some of the best-known brands in the world, and those brands are used by 2.5 billion people every day contributing to Unilever's purpose to make sustainable living commonplace. The website from Unilever (w.d.) states that all these brands are responsible for a turnover of 51 billion euros in 2018.

This research is done for Unilever Benelux, which is the largest fast-moving consumer goods (FMCG) employer within the Benelux. In the Benelux Unilever sells around 40 well-known brands like Axe, Lipton, Conimex and Calvé (see Appendix A). The turnover of Unilever Benelux in 2018 is around 1.5 billion euros. The emphasis is placed on the Dutch market and thus the Dutch part of Unilever Benelux. Unilever is in the Netherlands split in 4 product categories and 4 customer teams (see Figure 2). This decision will be clarified in Section 1.6.

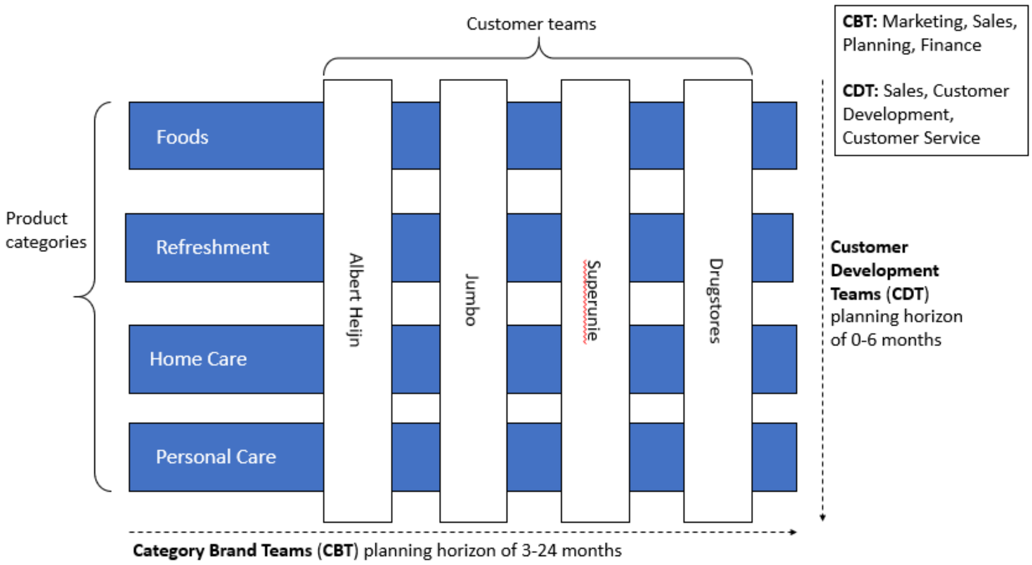


Figure 2: Organization matrix of customer teams and product categories at Unilever Netherlands - van der Poel (2010)

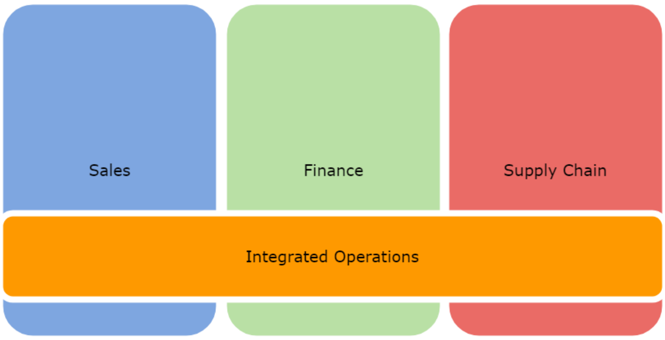


Figure 3: The position of Integrated Operations within Unilever

This research is conducted on behalf of the department Integrated Operations, see Figure 3 for the position within Unilever. Integrated Operations helps the business to run smoothly and involves all functions working together to create a strategic business plan which is then used to focus day-to-day actions. It identifies and drives demand, delivers customer service, optimizes Unilever's supply chain and generates demand volume and financial forecasts. Integrated Operations would like to implement a predictive model to generate reliable and automated forecasts for the total demand volume planning as part of the continuous improvement of the Sales and Operations planning (S&OP).

Concluding, in this section the organization structure has been explained to enhance the understandability. The retailers, departments and product categories will be anonymized before they are used in the further research*. Hence, the reader is able to position this research within the Unilever organization.

*Values in this thesis are manipulated for confidentiality reasons

1.2 Promotional demand forecasting at Unilever

Before continuing to more detailed information about the research topic, a general overview of the total demand volume forecasting process is given, see Figure 4. The demand forecast is split into the demand volume forecast and the demand value forecast. The first one is used as input for the Supply Chain department. The latter, is used as input factor for the value forecast of the Finance department. The focus of this research is on the demand volume forecast. This decision will be clarified in Section 1.6. The demand volume forecast is the input driver for the upstream supply chain. The demand volume forecast consists of the baseline and a lift factor.

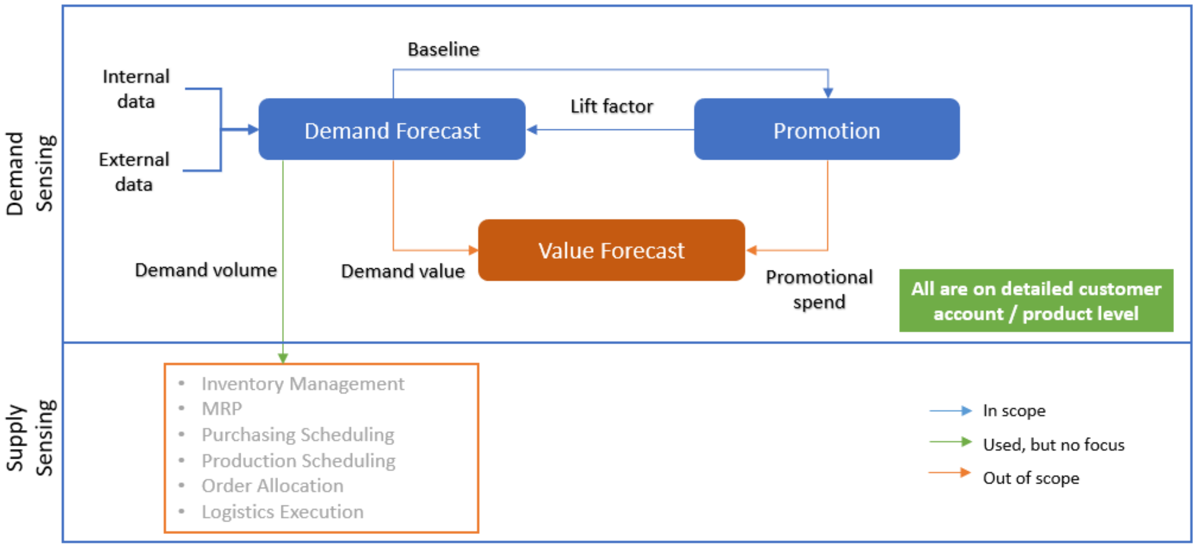


Figure 4: An overview of the demand forecasting process

The baseline demand volume is the expected number of products to be sold in a regular period of time without promotions. This baseline is calculated using statistical methods based on both internal and external factors. These statistical methods are able to determine seasonality in demand. A lack of external data makes it possible to forecast, but only less accurate.

The impact of promotions is described in a lift factor. The lift factor is calculated using machine learning based on the characteristics of the promotion. Because the baseline demand and the lift factor are critical factors in order to forecast the volume demand, they will be described extensively in Section 2.1.

All calculations and information within the demand forecast are on detailed customer account and product level. Therefore, when analyzing this information, the level of detail of the information can be adjusted to the reviewer' desire. The reviewer can be from the Sales, Finance or Supply Chain department. For example, the reviewer is able to look at aggregated information of all products sold at a particular customer, e.g. all products sold at Retailer A, or at detailed information of a particular product category (e.g. Foods).

1.3 Structure of this report

As stated in Section 1.1, this research is part of continuous improvement within the company. Inspired by The Lean Startup (Ries, 2011) fundamental build-measure-learn method, this section explains the structure of this report, which is illustrated in Figure 5. According to the theory by Ries (2011), the method consists of three parts: build, measure, and learn. In this research this theory is applied by combining both the current system (Chapter 2) and related literature to lay the groundwork of this research. Prior to this building step, the improvement cycle starts with motivating why these problems occurs now (Section 1.4), and clear problem definition is given (Section 1.5). Subsequently, these problems are analyzed (Section 2.6) to determine which problem will have the highest impact on improving the acceptance of the forecasted demand volume (Section 4.1). Thereafter, possible improvements regarding the problem with the largest improvement potential are drafted and analyzed (Section 4.2-4.5) to measure the impact on the performance of the forecast (Chapter 5). Lastly, the entire process is evaluated to learn whether or not to pivot or to persevere. The structure of this report is illustrated in Figure 5.

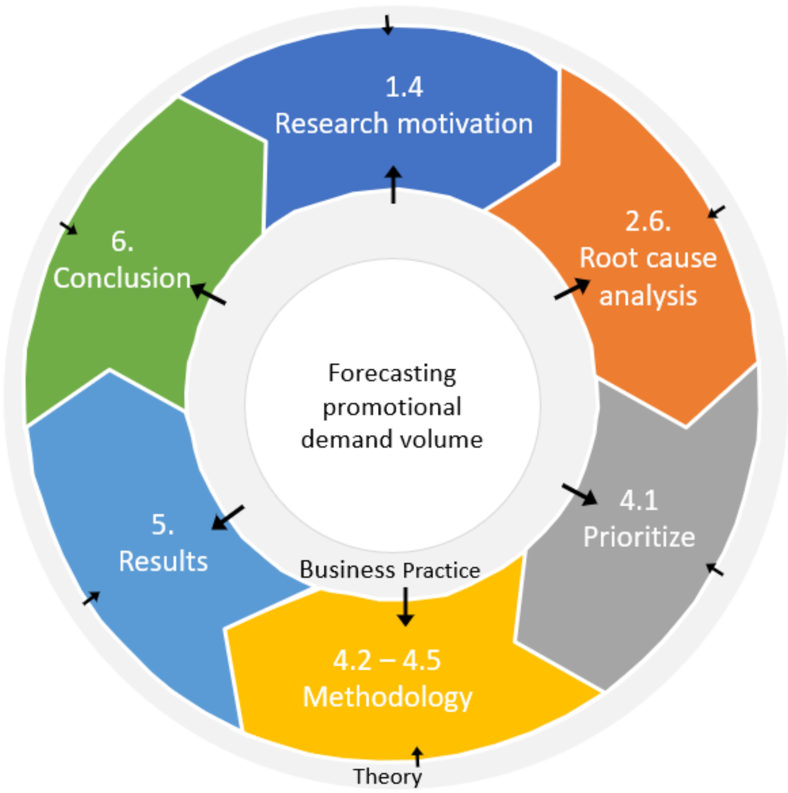


Figure 5: Structure of this report, inspired by the continuous improvement cycle

1.4 Research motivation

In the last decades, the promotional share of the total demand volume of products within the FMCG (Fast Moving Consumer Goods) market has barely increased, according to the external marketing database Nielsen (Nielsen ,w.d.). However, multiple price wars over the last couple of years increased the promotion pressure, to around 33 percent promotional volume of the total volume in 2018 (see Figure 5). Within the FMCG market competition is mainly focused on maximizing product availability at the lowest costs (Corsten & Gruen, 2003; van der Poel, 2010). In order to still be able to grow the business Unilever has to come up with a solution.

Therefore, Unilever wants to deliver the right amount of products at the right time to the right customer at the lowest cost possible. Prerequisite for this is an accurate forecast of the demand volume and value. The forecast accuracy has impact on the level of stock, stock costs and the service level, the factors that influence the product availability. On the one hand, a low service level because of under forecasting results in out of stocks in retailer stores and this affects the sales quantity and the relationship with the retailers. On the other hand, over forecasting results in extra stock, which leads to extra stock costs and potential obsoletes. Moreover, a high accuracy forecast leads to good business performance. Also due to the higher promotion pressure, the demand becomes instable. Because of this, Unilever noticed that their promotion forecasting became increasingly important.

The importance of product promotions for Unilever is endorsed by the promotion pressure (the percentage promotion demand volume of the total demand volume) of branded products in the different categories at the different retailers in the Dutch market (Derks, 2015). In Figure 6 the percentage promotions of the total volume demand is shown per cluster, as an average of all underlying retailers, to stress the importance of an accurate volume forecast for promotions for each cluster.

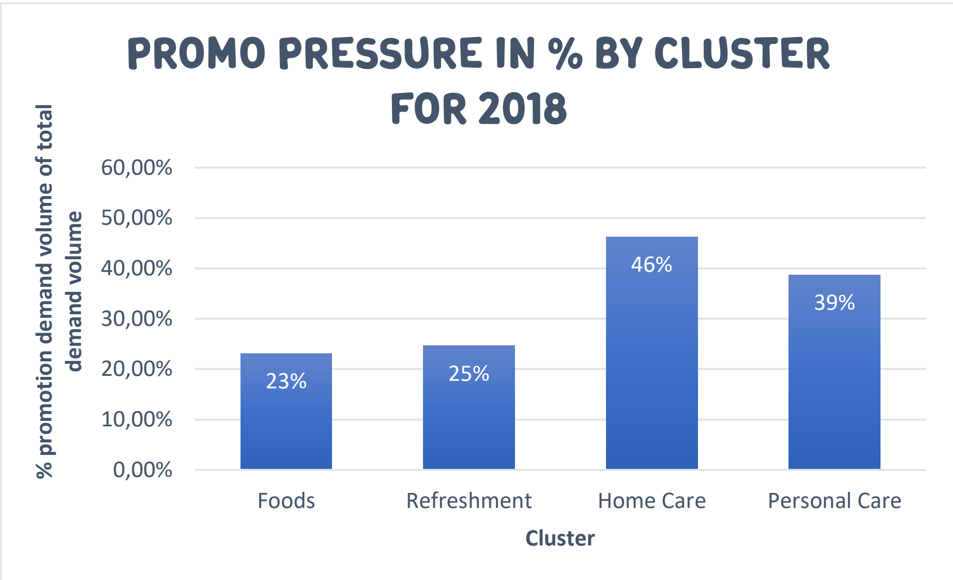


Figure 6: Promo pressure in % by cluster for 2018.

The importance of an accurate volume forecast for promotions has been the reason why Integrated Operations started in 2017 (Unilever) with the shift towards automating the demand volume forecast using a predictive modelling tool. Currently, in the implementation phase of the tool, Integrated Operations encounters resistance from employees against this tool to forecast the volume of promotions. Regardless whether the model might perform well on average at aggregated level (per account / per year) users do not trust the outcome of the predictive model because it does not perform well at their perspective, the detailed item level. This detailed item level is referred to as stock keeping unit (SKU).

Problems are faced with the use of the predictive model because of the discrepancy between the aggregated and detailed level performance of the predictive model which leads to the tool not being used. This results in low performance on the supply chain Key Performance Indicators: forecast accuracy and forecast bias. Also, this results in high stock costs or low fill rate due to under or over forecasting. According to the management the problem is considered to be on detailed level, because on aggregated level the average performance is considered good. Therefore, this research aims to improve the Unilever's current forecast performance at detailed level and focuses on model acceptance by the users.

Summarizing, the goal of this research is to achieve a model for the demand volume forecast of product promotions which is accepted by the users. Accepted in this sentence means that users trust the outcome of the predictive model and therefore fully use the model to forecast the volume of promotions and not have workarounds.

1.5 Problem definition and research questions

This section identifies the problem given the distrust of users in the predictive modelling tool. The Managerial Problem-Solving Method (MPSM) by Heerkens & van Winden (2017) is used to address the problem. To start with this method, the problem is defined based on the research motivation.

Problem definition:

Volume forecast for product promotions are not accepted by users and therefore not used.

The perspective of the total demand forecast, both regular and promotions, performance is two-fold. The management is interested in a good overall aggregated, based on the highest product and retailer hierarchy level, performance. On the other hand, users focus on detailed level, looking at each individual stock keeping unit (hereinafter shortened as SKU) at the lowest product and retailer hierarchy level. This research focus is on forecasting promotional demand. Therefore, from this moment on all information about forecasting refers to promotional demand forecasting, unless stated otherwise. The promotion demand volume forecast can be divided in two levels: macro and micro. Macro level is the top-down perspective where the forecast performance is measured on average and aggregated over all underlying promotions per year, per retailer and at the highest product hierarchy level. By way of contrast, micro level is considered to be the bottom-up perspective where the performance is measured at detailed level for each individual SKU, product group and retailer at the lowest hierarchy level. The performance on macro level is said to be good, according to the management, based on internal research at Unilever. However, looking at the performance from a micro level perspective there is often a discrepancy between expected demand volume by the users and the demand volume forecast by the predictive model. When the accuracy of the demand volume forecast of these individual SKUs is low this frustrates users, since they are judged by the performance of the individual SKUs within their customer or product group. Typical users are the Mid Term Planner (hereinafter shortened as MTP) and the Commercial Assistant Manager (hereinafter shortened as CAM). Due to these frustrations, users do not trust the output of the entire model and use their own alternative working method (e.g. Excel files). These workarounds do not contribute to Integrated Operations' desired situation of a data driven, one version of the truth and fully integrated volume demand forecast.

Figure 7 illustrates the trade-off between the applicability and accuracy between the two different hierarchy levels; detailed at SKU level or aggregated on lower hierarchy (e.g. product groups). This trade-off implies the difference of interest between the more accurate, but perhaps less useful, aggregated business performance at macro level on one hand and the more useful, but less accurate, forecast on micro level in detail on the other hand. The focus of this research is on the tension field between the performance of the model on macro level and the acceptance on micro level.

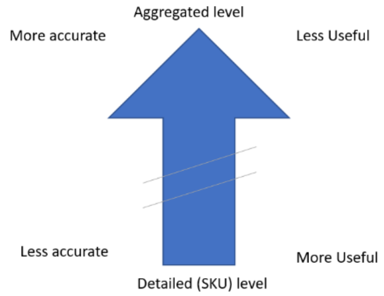


Figure 7: Trade-off between usefulness and accuracy based on level of detail.

Therefore, the main question of this research is:

Main question:
"How to improve the current forecasting method for product promotions to ensure it is accepted and internalized by the users?"

To help find the answer to the main question the following sub questions are drafted per area:

Forecasting method

The goal of this area is to obtain an excellent performing predictive model for the forecast of promotions.

Research questions:

Section 3.2: (A) What is the best forecasting method to forecast product promotions for a company like Unilever[1]?

Section 2.6: (B) Which factors impact the forecast method?;

Section 4.2: (C) How can these factors be enhanced to improve forecasts?

Sub question:

Section 3.3.4: (D) How to measure the performance of the forecast model?

Use and acceptance

The goal of this area is to increase the user's trust on the model, increase reliability of the results and get users to use the forecasting model. All this has a beneficial impact on the model performance. Acceptance by the users is requisite for using the model. Creating this acceptance is a continuous loop between increasing the model applicability and gain users trust to ensure the model is used correctly.

Research question:

Section 4.2: (E) How to increase the model applicability and acceptance by the users?

Sub questions:

Section 5.5: (F) How to ensure that the tool is used and internalized?

Data input

The goal of this area is to obtain reliable and relevant data input for the forecasting model.

Research question:

Section 5.2: *(G) What are the most important input factors for the model?*

[1] *with similar specifications as Unilever: FMCG, cross country category business*

1.6 Scope limitation

This section clarifies the decisions made in this thesis in order to maintain a structured and well defined research.

Country selection

Unilever Benelux consists of Unilever Belgium and Unilever Netherlands. In order to check whether these two markets can be aggregated to generate more data points a brief analysis is conducted on the comparability between the Dutch and Belgium markets. Unfortunately, the Belgium market differs too much from the Dutch market based on several aspects. A substantial part of the Belgium promotion market comprises of coupon promotions, while this type of promotion is rare in the Dutch market. Next to this, most promotions within the Belgium market are promoted on special displays and multiple items of a SKU are bundled together in a repack, sometimes even different SKU's are bundled together in one repack. These factors are likely to cause a difference in promotions mechanism within the Belgium and Dutch market. Because of this difference the model will not benefit from a larger set of data points. Since the research is conducted from the office in Rotterdam, the Dutch market is chosen for this research to make data collection easier.

Retailer selection

Incorporation of all retailers in the research will lead to extensive data gathering and might decrease the quality of the analysis. Therefore, four retailers are selected based on the following criteria: promotion pressure, size of the retailer and the data availability (see Section 4.3). Promotion pressure is the percentage of the total demand volume of a retailer that arises from promotions. The size of the retailer based on the total demand volume is compared to the other retailers to indicate the importance of a retailer. The criteria promotion pressure and size of the retailer are important also because the larger they are, the more you can say about the promotions in absolute terms, since it will reflect a larger portion of the total promotional demand. In order to use the predictive model data is needed. The data availability differs for each retailers, therefore this is also one on the selection criteria.

Forecast Perspective

The aim of this study is to bridge the gap between the functionalities of the predictive model, what users define as black box, and the usability by users. Building on an existing model, the aim is to increase the forecast performance on micro level in order to convince users to use and internalize the model, thereby accepting the promotion forecasts.

The responsibility for the financial forecast is not within the Integrated Operations nor the Supply Chain department. Therefore, the focus of this research is on the demand volume rather than the demand value. The financial data and parameters, that define the profitability of product promotions, are assumed to be correct and used in the way they are. This research takes a supply chain point of view by focusing on demand volume. Therefore, the performance of the forecast is measured based on the key performance indicators (KPIs), forecast bias and forecast accuracy, which are used globally within each supply chain department of Unilever.

Business perspective

Syntetos, Babai, Boylan, Kolassa, & Nikolopoulos (2016) distinguishes the supply chain into manufacturer, retailer and consumer based on several dimension factors like location, timing and product aggregation. All forecasts within this research relate to the demand forecast of shopper behavior, see blue circle in Figure 8. Shopper behavior relates to the understanding of how consumers behave as shoppers at the different retailers, which is in this thesis also referred to as In-Market quantity. To illustrate, when forecasting the expected demand quantity in a promotion we look at the expected amount of products bought by consumers at the retailer. Next to this In-Market perspective there is the Ex-Factory and the Phasing perspective (see Figure 8) which are both not in scope for this research. The Ex-Factory perspective comprises the relation between Unilever as a manufacturer and the retailers. The Ex-Factory quantity is related to the predicted retailer behavior, the amount of products that are ordered by the retailer in order to supply their stores with products to sell to the consumers. The Phasing perspective covers the logistics part of the process, meaning to predict what amount of products needs to be delivered at which retailer at what exact week in time in order to have sufficient stock at the start but also during a promotion. Whereas in the In-Market perspective the focus is on forecasting the demand volume for an entire promotion. Promotions might last longer than one week.

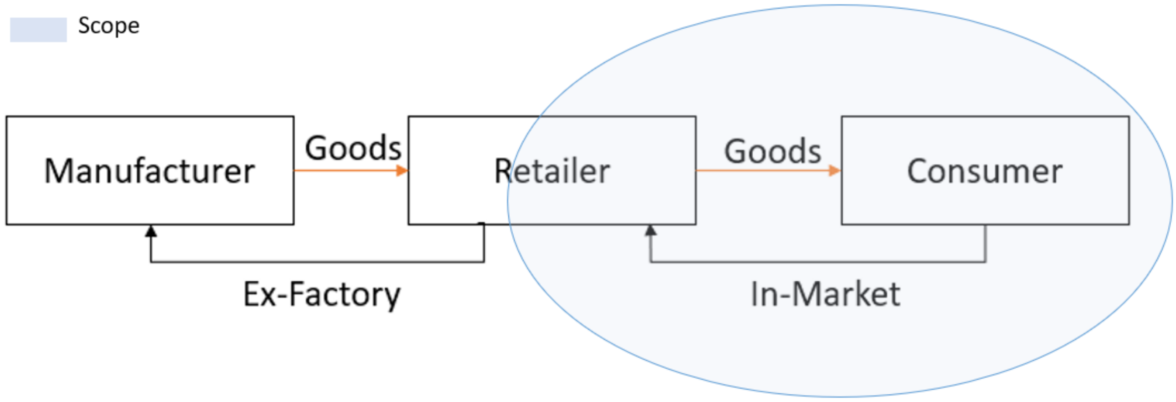


Figure 8: Interaction between In-Market and Ex-Factory demand volume – (Syntetos et al., 2016)

Organizational perspective

The demand volume forecast for product promotions impacts the whole supply chain. However, from an organizational perspective the primary focus of this research is on the users working with the tool: CAMs, MTP, and the external company (VisualFabriq™). These users, CAM and MTP, are easy to approach, use the tool daily and have the most interest in a well-applicable model. However, the perspective in this research is from the users that are directly connected to using the predictive modelling tool for the volume forecast of product promotions. Secondary stakeholder is the Management. Management in this case is used as a collective term of all people involved with running the business on a strategic level. This is why from an organizational perspective the focus is on the users. Although the management will eventually also benefit from this research.

Data analysis and software usage

Both qualitative and quantitative data analysis will be conducted in order to answer the research questions. Qualitative data will mainly be obtained by semi-structured interviews with CAMs and MTPs to define their issues with using the tool and gather feedback for the improvement of the model applicability. Quantitative analysis will be performed in order to quantify the current situation, to select performance measurement and to find the right forecasting model that fit the companies requirements.

Data input is briefly explained because it is a requirement for a well-performing model. The master data quality is out of scope, because it cannot be influenced in the limited amount of time available for this research. To support this research the software Python, JupyterLab, Minitab, Microsoft Excel and PowerBI will be used.

Practical requirements

Besides the scientific nature of this research, practical goals should be defined as well. The goals is to improve the applicability of a good performing forecasting model for Unilever Netherlands. Two sub goals are given to reach this goal.

- Ease of use
  The forecasting model should be simple to use. Thus, a Unilever employee should not have to put much effort in using the model. This means that the improved interface of the tool should be clear and simple. Also, the output of the model should be easy to interpret. Semi-structured interviews within Unilever indicated that the results generated by the model should be understandable as well as the model itself, in order to increase the acceptance of the forecast of the model.
- Building on existing models
  The predictive model should build on the already developed tools and should work with data readily available for the users. In order to enhance the direct applicability in the current way of working.

2. CURRENT SYSTEM ANALYSIS

The aim of this chapter is to describe the current prediction process, the model and tool to forecast the promotion demand volume. It explains what a product promotion at Unilever comprises. The way of working at the start of this research is described as the initial situation in order to make a comparison with findings of this research described in Chapter 5.

First, an overview of the relations between the forecasting model, data input and the user and its acceptance is given to enhance the readers understandability (see Figure 9). Essential for a good model is reliable data as input. In this case the data is partly entered by the user and partly loaded from multiple databases. Proper use and reliable data are key for the performance of the model. If the performance of the model matches their expectation users will accept the output of the model and thus use the model to predict the demand volume of promotions. Decent use of the model leads to more reliable data and more reliable model output. This loop of acceptance by the users will eventually result in an applicable and well performing model. These relations are illustrated in Figure 9.

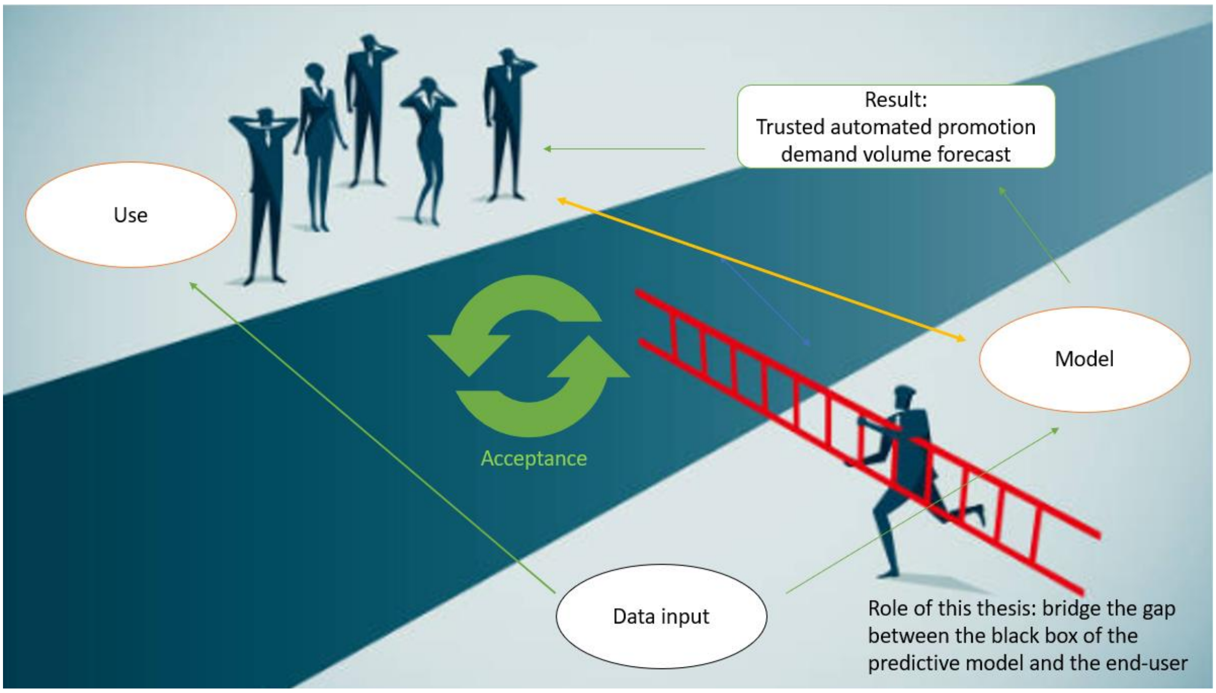


Figure 9: Illustration of the perception of the forecasting model at the starting point of the this thesis with the possible relations.

2.1 Product promotion at Unilever

Unilever uses different types of promotions. E.g. the well-known buy one get one free ('1+1 gratis') during the '*special promotion weeks'* at Retailer A or other retailers. Almost all promotions are price promotions; the consumer gets a reduced price in one or another form. But besides price promotions occasionally a promotion is in the form of a coupon or with a premiate or free product (e.g. discount for a theme park, free sample of a (new) product). The success of the different type of promotions is influenced by numerous variables (Van der Poel, 2010). Because of the high dependency on external factors only price promotions are taken into account in the model. These price promotions are described by Unilever as 'Regular and Category promotions' and include the mechanisms % discount, buy one get one free, single price discount, and buy X for Y. In a desired situation all of these variables are taken into account in the predictive model to forecast the promotional demand volume

The promotion mechanism is one of the characteristics of a promotion. Next to this, the CAM fills out several other parameters (see Figure 13 in Section 2.4) like number of SKU's to include, whether the promotion will be displayed in the folder or gets extra space in the retailer store (2nd placement), in order to define a promotion. Based on these parameters and the available historical data the predictive model calculates the lift factor. Basically, the promotional demand volume is determined by the baseline demand multiplied by the lift factor, see formula 2-1.

$$Demand\ volume\ =\ Baseline\ volume\ *\ lift\ factor$$ **Formula 2-1**

The baseline is the amount of products that are expected to be sold in a regular period of time without promotions. This baseline is forecasted based on statistics and actual demand in the past, the responsibility to check whether this amount of baseline makes sense is at the MTP. For example, when a product is seasonal (e.g. Unox, which has a high sale in the winter) the MTP has to check whether the model determines a corresponding seasonal pattern is the forecast (relative low baseline in the summer period and high in the winter). In this research the method of forecasting the baseline with regular demand is assumed to be good and not further investigated.

The lift factor can be seen as the expected increase in sales when a product is promoted. Theoretically this lift factor can be negative, but in practice this will never be the case. Because when the expected sales promoted product would be lower than in regular sales, the company simply decides not to promote that product.

Next to the baseline volume and lift factor also cannibalization and market intelligence impact the total demand volume. The cannibalization effect is the effect promotions (P) have on the baseline, see Figure 10 where the promotion demand volume fully substitutes the baseline demand resulting in a dip (D) in the period after the promotion. In this matter market intelligence is the sales effects that Unilever expects on top of the promotion and demand volume (e.g. expected annual growth taken into account when forecasting the volumes).

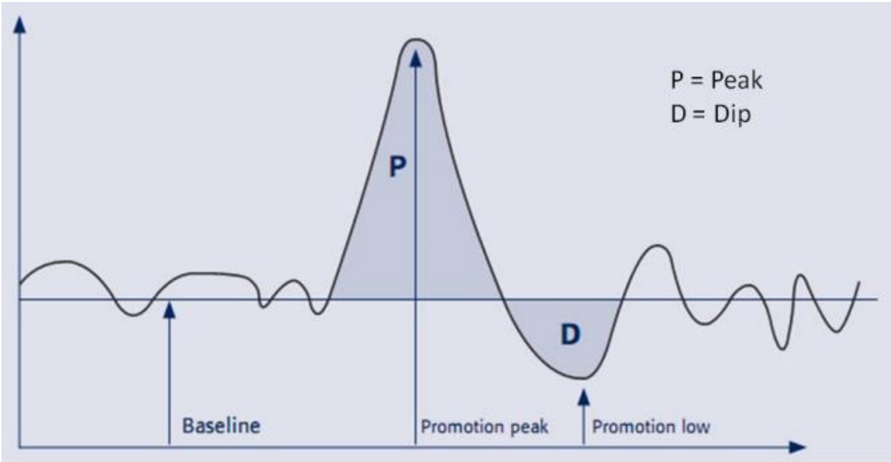


Figure 10: Impact of a promotion on the demand volume. Source: www.eyeon.nl

2.2 Promotion planning process

The promotion planning process steps are described in this section and is defined by the processes steps shown in Figure 11. Starting point is the promotion year plan which contains the promotion frequency for each product category, the year plan is determined at the end of the previous year. Within this interval pattern the account manager can determine the content of the promotions. The content of the promotion is characterized by its parameters, these parameters will be explained in Section 4.3.

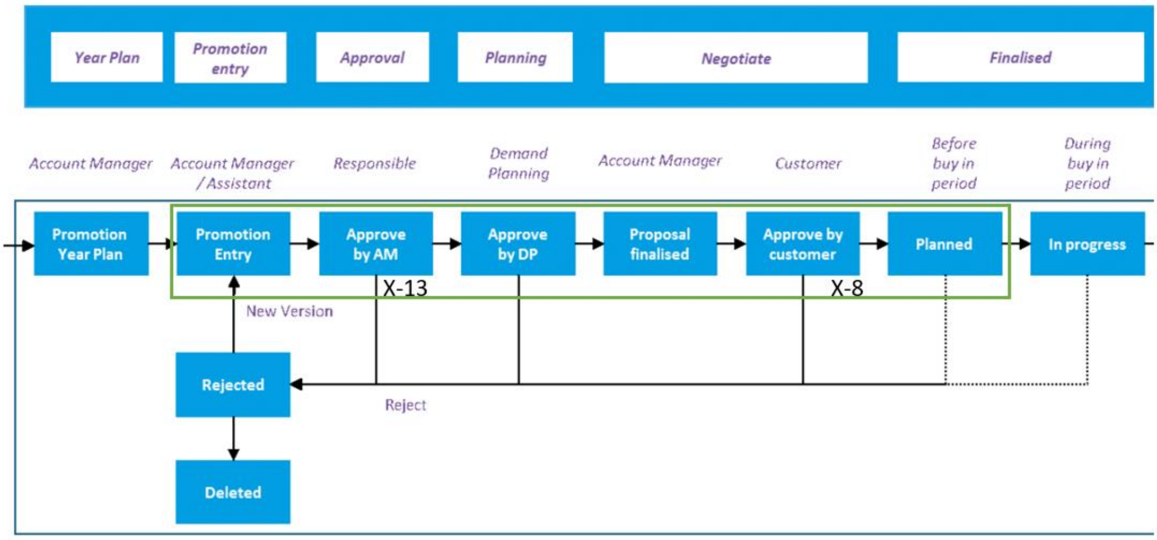


Figure 11: Process flow of product promotions. Source: Visualfabriq Trade Promotion Master™

In general 13 weeks before the starting date of a promotion the CAM and the retailer discuss which products will be promoted, the CAM proposes the expected demand volume of those products during that promotion. The so called promotion proposal or 'Actievoorstel'. This timing is internally referred to as X-13. In order to propose these demand volumes to the retailer the CAM predicts the demand volume of products often referring to volume of last similar promotions. Semi-structured interviews with the CAMs indicated that they often compare promotions in Excel by loading the data of previous promotions into a datasheet in order to manually use the VLOOKUP function in Excel to be able to compare the demand volumes of the same products from different promotions with each other. Based on this comparison and personal experience the expected demand volume are proposed. At X-8 the retailer, confusingly called customer (not the consumer), sends a promotion confirmation also known as 'Actiebevestiging' and at X-4 the retailer sends an updated version of this 'Actiebevestiging' as a confirmation. It might happen that in this stage the assortment within the promotion is slightly changed or proposed volumes are updated according to new information or personal judgement. For example, a change in the position a product gets in the store of the retailer or change of product within the promotions. The CAM has to react to these changes because it might have significant impact on the supply chain, for example a recently added product might be out of stock and not able to be delivered on such short notice. To summarize, for promotions there are three moments in time, X-13, X-8 and X-4, at which the demand volume is forecasted. Over time adjustments are made based on updated user or retailer information. However, sometimes these adjustments are made when there is no more time for the supply chain to react on these changes due to for example large product lead times. Hence, an accurate forecast at an earlier point in the process is desired to make it possible to adapt to these adjustments further upstream in the supply chain. When this earlier point in time is differs per cluster, retailer and products.

2.3 Current forecasting method

In cooperation with an external IT company the construction of the model was started in 2017 and is currently in the implementation phase. In this thesis the words 'tool' and 'predictive model' will be used for the model. This section briefly explains the current forecasting method in order to enhance the readers understandability.

The word tool is typically referring to the software environment and interface, developed by the external company in collaboration with Unilever, used by the users to for example fill out the parameters to describe a promotions characteristics. The model is able to forecast the demand volume based on several features using machine learning. The model is used to forecast the demand volume of promotions that will occur in the future, therefore it is called the predictive model. In this research the words predictive model and forecast model are interchangeable. Both the management and the users of the tool, define the operations of the model as a black box. Hence, a brief and elementary explanation of the predictive model is given for the ease of understanding.

The model is developed to predict the promo types that occur the most and show the most regular pattern. These promo types are 'Regular' and 'Category'. The characteristics of these two promo types have the best fit with the predictive model. Because of their low variability in demand pattern and sufficient data availability. Other types are not eligible to forecast with the predictive model at the moment. Therefore only the promo types 'Regular' and 'Category' are selected at this stage. Users enter the data of these promotion types in the tool using a user interface. Based on this input data the predictive model makes it calculations and forecasts the demand volume of future promotions. The user is able to make adjustments to this forecasted volume based on experience and personal judgement. When the volumes are confirmed by the retailer they will be communicated to the factories upstream the supply chain. Ideally, after the actual volumes, based on scanning data, are received from an external company, the promotion should be reviewed by the users to check whether the planned promotion variables (e.g. volume or promotion mechanism) corresponded with the actual promotion variables. However, according to the S&OP Lead Manager, the extent to which a review is actually carried out is limited. The reason for reviewing the promotions is the learn from the adjustments and to eliminate undesired situations, like stock outs or incidents at a factory, from the data in order to train the model on proper data instead of those occurrences.

The predictive model is based on a multiple linear regression analysis and is written in the programming language Python. An intermediate level of Python coding is required to understand how the model is operating exactly. Python has many libraries for data loading, visualization, statistics, and more. One of the main advantages of using Python is the ability to interact directly with the code, using a terminal or other tools like the Jupyter Notebook (Müller & Guido, 2017). The Jupyter Notebook is an interactive environment for running code in the browser. This environment is also used in the research for the analysis of the predictive model.

2.3.1 Performance measures at Unilever

The performance of the model output is measured using the Key Performance Indicators (KPIs) of the Supply Chain department within Unilever. These KPIs are forecast accuracy and forecast bias: The forecast accuracy is calculated by dividing the absolute difference at product level between the actual sales quantity and the forecast demand quantity by the actual sales quantity multiplied by hundred percent and subsequently subtracting all this from 1, see formula 2-2. The second part of this equation is in research referred to as the mean absolute percentage error (MAPE). The formula is as follows:

$$Forecast\ Accuracy\ \% \ = \ 1 - \frac{(ABS(Actual\ Sales\ quantity - Forecast\ demand\ quantity)}{Actual\ Sales\ quantity} * \ 100\%$$

Formula-2-2

The Forecast Bias is calculated (formula 2-3) by subtracting the forecast demand quantity from the actual sales quantity and dividing this by the forecast demand quantity and multiply this all by hundred percent. The formula is as follows:

$$Forecast\ Bias\ \% \ = \frac{(Actual\ Sales\ quantity - Forecast\ demand\ quantity)}{Forecast\ demand\ quantity} * 100\%$$

Formula 2-3

The forecast bias could be either positive, negative or zero. To calculate the forecast bias of a promotion the forecast bias of all underlying products is summed. In this way it might occur that a large positive bias and a large negative bias cancel each other out. A negative bias corresponds to the an over forecast, when the forecast demand quantity is larger than the actual sales quantity. A positive bias corresponds with an under forecast, when the forecast demand quantity is lower than the actual sales quantity.

2.4 Input data

The input data for the predictive model consist of two parts. First part is the internal and external data that are used by the model. The second part of input data is all the information provided by the users.

The first part, internal- and external data(bases) are used to be able to compare promotions with historical data and actuals. For example, scanning data from shopping behavior from consumers is bought from the external marketing database Nielsen to see what is actually sold by the retailers. Based on these actuals the performance of the predictive model is measured. In the current model, there is no automatic control whether a file is loaded or not. At the moment, one employee might have substantive knowledge about these data loads and performs these uploads manually. Absence of this employee or wrong timing affects the performance of the model. In that case it might occur that the comparison between the number of products predicted and the actuals are based on different timings. These dependencies create a need for automation of these data loads to ensure reliability.

The second part is a result of the information that the users filling parameters in the tool, see Figure 13 in Section 2.5. An example is the promo type, for example two products for the price of one. However, users lack knowledge regarding the impact of such parameters on the output of the predictive model. This lack of knowledge contributes to the perception that users see the model as a black box of which they have no understanding how it actually works. To take a step forward from this point three years of historical data is cleaned during this study, meaning checked to make sure all parameters are filled in correctly, in order to provide training data for the predictive model.

Master data, containing information of products like productid and product dimensions (e.g. the number of products in a case), is regularly updated because there are multiple product changes/ innovations in a period of time, one of the characteristics of a FMCG. It is essential to link innovations of a product to the correct product hierarchy level in order to be able to connect comparable historical sales patterns to the forecast of this new product. A product with a small change, for example a package change for washing powder, is assumed to have the same sales pattern as the same product in the previous package. Currently, these frequent changes in master data cause many errors in the forecast output. If a product history is not correctly connected the model considers this period as a stock out period, resulting in a large deviation from actual demand.

Together the section 2.3 Current forecasting method and 2.4 Input data form the basis of the predictive modelling tool. In Section 2.5 describes the users and the way they interact with this tool.

2.5 Use and acceptance

Three types of users are defined in this research. The Mid Term Planner (hereinafter shortened as MTP), the Commercial Assistant Manager (hereinafter shortened as CAM) and the 'experts on the predictive model from Unilever' (hereinafter referred to as super-users). The super-users are within the Integrated Operations department and have an admin role. The MTP is responsible for a correct baseline demand planning. The CAM is responsible for the promotion proposals and fills out the parameters (e.g. promo mechanism or number of products) that identify a promotion in the tool, see Figure 13.

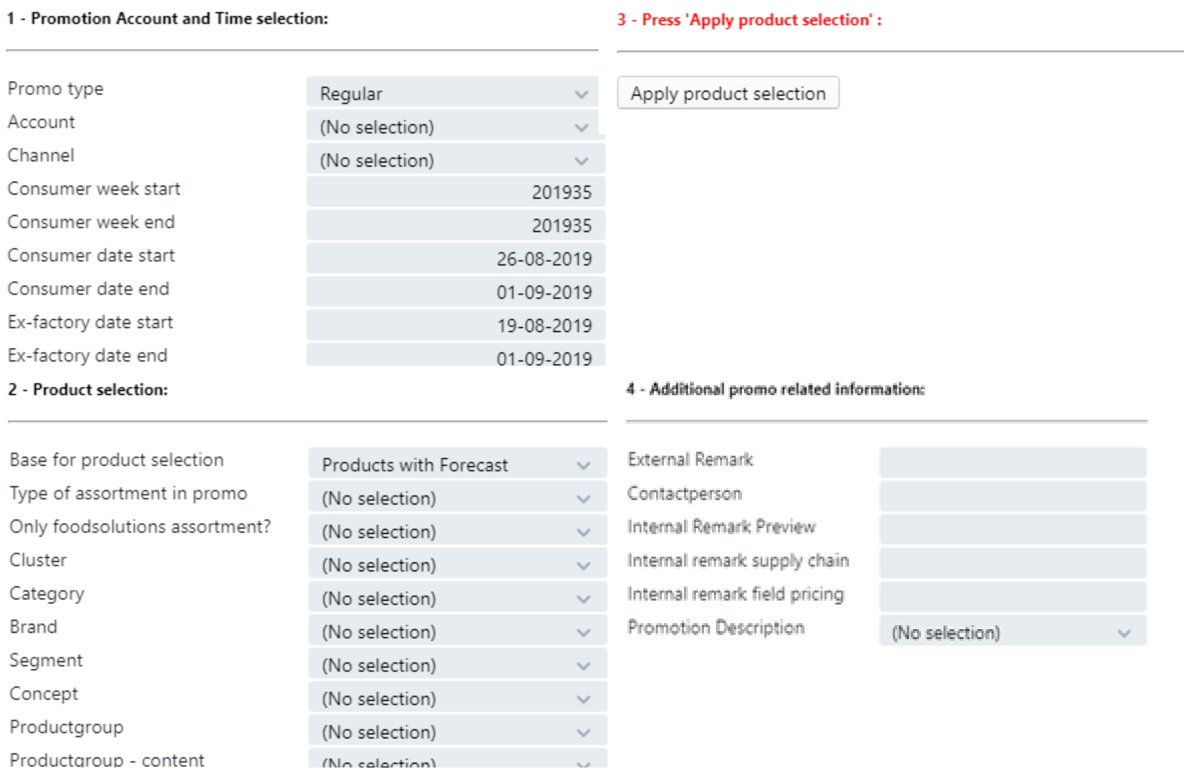


Figure 12: User interface to enter the input parameters for promotions

When one encounters an issue it escalates it to the super-user without much effort in trying to find the cause of the issue. Super-users try to solve the issues that are raised by the users and focus on continuously improving the tool.

The CAM focuses on the output of the forecast model. In a way that the output is compared with historical data. This comparison is often done in a personal Excel file, personal expectations of the CAM, and retailers expectation. When the model output is obviously wrong, according to the CAM judgement, they will override this suggested output. When the CAM often has the feeling they should override the suggested output they will start to distrust the capabilities of the model. Therefore, a CAM will rather use his/her personal way of working (e.g. separate Excel files).

Besides this, it is not only the output of the model that causes distrust also technical- and user issues cause distrust in using the model. Technical issues occur because the tool is still in the implementation phase and further developments are made continuously. CAM lack knowledge regarding the impact of parameters on the predictive model. Both MTP and CAM lack time and knowledge to be able to find the cause of the issues they raise towards super-users. Often, super-users have to spend a lot of time solving each individual issue because the CAM or MTP provides them with little guiding information. When encountering issues the CAM and MTP escalate this to the super-users. Therefore, super-users are busy trying to fix these issues reactively. While, the super-users should rather spend their time on fixing root causes and further improving the tool. These issues by the users can be divided into two groups. The first group are issues that are reported by the user to the super-users. The second group of issues are not reported by the user to the super-users, however users use a work-around for the process because they encounter some kind of issues with using the tool. For example, when the assortment of products in the tool is incomplete or when users have a bug while filling out the parameters.

At the moment, the predictive model is not fully incorporated in the way of working because users encounter errors working with the tool and the performance is low on micro level. As a consequence users work with their own workaround (e.g. Microsoft Excel). There is a mismatch in what the user expects the model to deliver and what it actually delivers is because there are a lot of exceptions which do not fit the model, yet. What these exceptions exactly are and how often they occur is not yet fully documented at this moment. It is also not known how to deal with these exceptions.

To conclude, as a result of the distrust of users in the model output they make manual adjustments based on their experience and knowledge. However, in order for the management to make the right decisions to achieve their targets the forecast needs to be true and free of biased behaviors and assumptions. Therefore, only by being truthful and honest in the forecasting practices can we really steer the business and achieve all our targets. The business must start to demand the best forecast no matter the provenance. In Chapter 3 we try to find the answers in literature to the sub questions from Section 1.5 given the current situation at the areas Model and Use explained in Chapter 2. In order to test alternative forecasting models in Chapter 5.

2.6 Root Cause Analysis

The goal of this analysis is to investigate the problem areas that have an impact on the promotional demand volume forecast and to define the focus of this research. Users distrust the volume forecast for promotions. Therefore, users do not accept the volume forecast of the predictive model. If users do not trust the outcome they will use a work-around which leads to a decrease in reliability of the input data for the model. Reliable data input and good use of the tool are essential for a good forecast accuracy. The forecast accuracy impacts the level of obsoletes, stock costs and service level. To give an example, if the volume forecast is too low, products will run out of stock, which affect the service level. And, if the volume forecast is too high, the surplus of products in stock will be higher which leads to higher stock costs and potential obsoletes. This stresses the importance of an accurate forecast.

In order to relate the research topics (sections 2.3-2.5) to the business context the topics data input, use and acceptance need some clarification. According to Kim, Byoung-Ju, Euj-Kyung, & Doheon (2001), a major problem is that data in data sources are often 'dirty'. If a high proportion of data is dirty, this will surely result in a unreliable forecasting model. Dirty data is usually presented in the three forms: missing data and wrong (noisy) data, and non-standard representations of the same data. For the remainder of this research, the missing data and non-standard representation of the same data is represented by data availability, and wrong (noisy) data is interpreted as data quality. The research topic use and acceptance is clarified into business context by terms recognizable to the company: retailer dependency and internal processes.

Retailer dependency

It is up to the retailers how much and which information they share and in practice it is common that not all information (e.g. stock levels) is shared, mainly because of data sensitivity. Also, promotions at other retailers can result in last minute changes to the retailers promotions when the discount percentage at the competitor is higher.

Forward buy of retailers, the factor of additional products compared to actual sales in the promotion period, retailers order against promotional price, increases the complexity of the forecast. Because it leads to extra costs and less efficient promotions as the discount is given to retailer and not to the shopper. This makes the forecast harder for interpretation. This is one of the reasons that the focus of this research is on the In-Market perspective, looking at the actual scanning data of consumers.

Internal processes

Multiple teams are divided over the customers and are all working on forecast product promotions of their customer. There is no standard way of working and since information sharing between the customer teams is limited it evolves in multiple ways of working. The process flow (see Figure 11) is not always followed resulting in the right data is not available because it was not updated at the right time. Sometimes a CAM makes adjustments to the parameters of a promotion to reach a volume target or in consultation with the retailer. These adjustments affect the quality of the input data of the forecast model because it adds one-off information that you do not want to take into account in future promotion forecasts. Also, the timing of these judgmental adjustments are crucial to the model performance. Promotion volume forecasting is only one of the tasks of a CAM. This results in the fact that, when facing an issue, immediate escalating an issue without taking time to tackle the issue self. In a FMCG environment frequent changes of products occur, thus code and products-ids switched and not yet processed. These switches often lead to issues in the current prediction system.

Data availability

In order input a promotion multiple data sources need to be consulted for each promotion. This is a time consuming and user unfriendly process, which does not contribute to the use of the tool and therefore the forecast accuracy is affected. When entering a promotion in the tool many information has to be filled in, for which the user does not know what the impact will be of those particular part of information or the lack of information on the forecast accuracy. Users encounter many issues while working with the tool, both technical and user related issues. Alongside, users lack knowledge about the use and the output of the predictive model.

Data quality

In the current forecasting process promotions are not or limited reviewed based on their actual performance. This review is essential to check whether both the forecasted volume and the promotions parameters were accurate compared to the actual demand. Data input is not always reliable, meaning that not the right data is available at the right time. Frequently mismatches in timing of the data occurs. There are no quality check build in the user interface to check if the entered information makes any sense.

Forecasting method

Until recently the forecasts are made by users based on experience without support from a mathematical model. With the implementation of the forecasting model and tool this forecasting process should become more automated. However, at the moment almost all promotions are still forecasted manually since the forecast generated by the model is not accepted by the users. The volume forecast is based on experience of the users rather than substantiated by data or statistics. This is why the current forecasting process is time consuming.

Conclusion of the root cause analysis

To conclude this section, the areas 'data availability', 'data quality', and 'forecasting model' indicated in Figure 14 by the color blue are in scope for this research. Nevertheless, the other areas might also be improved by the outcome of this research but these problems are about organizational change which is hard to influence given the limited amount of time. It can be concluded that the current forecasting process is result oriented rather than process oriented, taking the time to find the cause and try to learn from it. If the predictive model tool is fully and correctly used througout the organization it would result in a data driven, one-version of the truth, regarding the volume forecast, which leaves less space for objective influence by the organization but saves a lot of time in the forecasting process. The next section will give a scientific background to this research by providing literature related to the problems. However, the time and the influence of this research is limited, Section 4.1 will therefore discuss the prioritization of these problems in order to determine which problem to solve first in order to gain the largest improvement on the forecast accuracy.

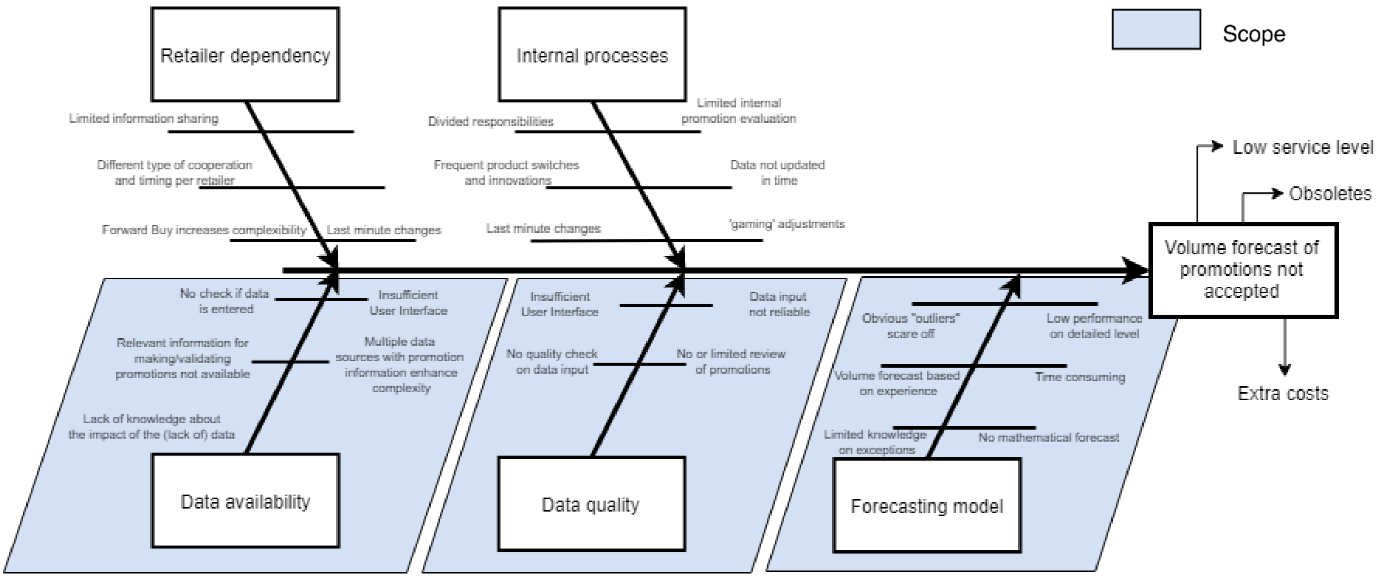


Figure 13: Root Cause Analysis of the current system

3. LITERATURE REVIEW

In this chapter, relevant scientific literature is discussed. The goal of this chapter is to get answers to the research questions from a scientific literature perspective. Therefore, Section 3.1 will give an introduction to the subject, where Section 3.2 will help to answer research question A: *'what is the best forecasting method to forecast product promotions for a company like Unilever?'*. Section 3.3 describes relevant scientific research that will be used in the methodology and analysis part of this study.

In more detail, this chapter starts with the definition of promotional forecasting in Section 3.1.1 Followed by the impact of accurate forecasts on the supply chain in Section 3.1.2. Section 3.1.3 describes the characteristics of a fast moving consumer goods company to give an idea about the environment of this study. Subsequently, Section 3.2.1 describes and elaborates the different type of forecasting methods relevant for this study. Based on related scientific research in Section 3.2.2 the forecast method for product promotions is selected. Section 3.2.3 gives more information on forecasting methods used by different fast moving consumer goods companies to answer research question A. Section 3.3.1 explains about the dependent and independent variables used in different studies. Section 3.3.2 explains two factors that need to be addressed when forecasting based on explanatory variables. Lastly, Section 3.3.3 describes which performance measures are commonly used in forecasting.

3.1 Literature on relevant topics

This section introduces the topics related to forecasting product promotions.

3.1.1 Promotional forecasting

Promotions, also referred to as trade promotions, include special pricing and sales incentives, discounted or free display fixtures, trade shows, demonstrations, and no-obligation gifts such as tickets to sporting events or novelties (pens, paperweights, calculators). Forecasting these trade promotions is called Trade Promotion Forecast (TPF) and refers to the process that seeks to discover correlations between trade promotion characteristics and historical demand, in order to arrive at an accurate demand forecasting for future promotion campaigns. Key to modelling promotion behavior is the ability to distinguish the increase in demand due to the impact of the trade promotion in contrast to the baseline demand without any promotions. The increase in demand due to the impact of trade promotions is called the lift factor.

Furthermore the lift factor as dependent variable can be transformed in multiple ways. There is no conclusive research on the performance of the different forms of the dependent variable (van der Poel, 2010). In this research lift factor represents the promotional demand volume divided by the baseline demand volume of a SKU. The advantage of using the lift factor as dependent variable in the model the promotional demand volume is standardized against the baseline volume. As a results, the absolute demand quantity height of a promotion has been removed from the predictive model equation.

An unwanted side effect of promotions might be cannibalization. Cannibalization is determined from sales data as the ratio between the volume drop of cannibalized product and the volume uplift of the promoted product (Herrala, 2018). For example, if product A is in promotion, the cannibalization effect occurs when customers buy product A instead of regular product C. Also, when customers buy product A in week X during promotion, instead of buying product A in a regular week Y. Van Donselaar, Van

Woensel, Broekmeulen, & Fransoo (2006) state that one of the success factors of a promotion is determined by the substitution effects (consumers switching between different products of the same category).

Besides evaluation of individual promotions, it is important to be aware of the fact that a higher number of promotions of a product positively affects the lift factor. Derks (2015) recommends to take the expected cannibalization of base demand into account and investigate the optimum number of promotions in order to maximize company profit.

Until the emergence of automated promotion planning methods, it was common practice for retail store managers to use the "last like" rule when ordering inventory for upcoming promotions. This means that they ordered the same quantity of products that was sold during a similar promotion in the past (Cooper, Baron, Levy, Swisher, & Gogs, 1999). Now, progress in technology offers better ways to handle the volume planning. Therefore, the reliance on the simple "last like" rule became inefficient (Trusov and Cooper, 2006)

3.1.2 Impact of an accurate forecast

This section describes the impact of an accurate forecast on the supply chain. Forecast accuracy can be described as the relative difference between a the forecasted number and the actual number. An increase in forecast accuracy will lead to a reduction in variability. A reduction of variability in consumer demand downstream the supply chain will decrease the bullwhip effect (Lee, Padmanabhan, & Whang, 1997). Therefore, lower safety stocks are necessary, resulting in less stock costs, less obsoletes, while maintaining a certain fill rate and customer relationship. The goal is to maximize the forecast accuracy while minimizing the total costs of amongst others executing the forecasting process. Kerkkanen (2008) states that knowing the role of forecasting and the impact of forecast errors create a basis for defining a realistic target for forecast accuracy, identifying the most important customers and products to be forecasted, and finding a suitable way to measure the forecasting performance. Next, including external information (e.g. market intelligence) can improve the demand forecast performance (Currie and Rowley, 2010) especially in fast changing environment. Like this research, most forecasting techniques and promotion models focus on forecasting future consumer demand. However, in his research Kerkkanen (2008) warns that there is a risk that unrealistic accuracy targets and fraudulent error measures are adopted if the environment is different. That is the reason why in the next section the forecasting environment of Fast Moving Consumer Goods is described.

3.1.3 Fast Moving Consumer Goods company characteristics

This section describes the characteristics of a Fast Moving Consumer Goods company from a supply chain perspective by focusing on the demand volume. Fast Moving Consumer Goods (FMCGs) are defined as products which are sold rapidly at relatively low costs. These products are necessities which a consumer buys within a short interval of time, without spending little of no effort on the purchase decision. According to Singh (2014), advertising and suggestions of friends and neighbors usually play a major role for trial of new FMCG products. In his research Adefulu (2015) states that, the heart of the FMCG business is the competition to attract consumers' attention towards products or services. The prominent tool for attracting consumers' attention towards products is product promotion (Chaharsoughi & Yasory, 2012). Therefore, the performance of promotion have major impact on the company's market share.

From a supply chain perspective the main characteristics of FMCGs are high volumes and high inventory turnover, which denotes the number of times inventory is sold in a period of time. FMCG competition is mainly focused on minimizing out of stock, therefore maximizing the product availability at the lowest costs which is of great importance during promotions. Achieving the highest product availability at the lowest costs in combination with high volumes and high inventory turnover is a big challenge. Because of this, Unilever strives to deliver the right amount of products at the right time to the right customer at the lowest cost, which is only possible with an accurate forecast of the demand volume.

The FMCG segment is highly dynamic and innovative. The FMCG companies are under pressure to keep innovating their products, either the changing content or the package, in order to keep being attractive to customers. For the company these rapid pace of innovations and product changes have a large impact on the product life-cycle management. All product switches and changes must be recorded in order to keep track of the product development over time. Only if product development over time is recorded properly its data will be useful for forecasting the product demand in future promotions. For instance, when forecasting the demand of product A, the demand of similar products is compared in order to come up with a decent forecast for product A. When the package of that product A is slightly changed you might still rather want to forecast its demand based on comparable products to product A instead of forecasting the demand of this 'new' product, without any historical data. The latter will be much harder to forecast, because there is no prior data to base the forecast on.

Consumer demand patterns of FMCG at different channels are driven by different factors (Shankar, Inman, Mantrala, Kelley, & Rizley, 2011). Therefore, it is important to develop separate demand forecasts by channel. Accurate demand forecasting in each channel is also critical to managing the business and expectations of managers (Shankar et al., 2011).

3.2 Literature for answering research questions

This section describes the literature that helps to the research questions.

3.2.1 Forecasting methods

Given the high inventory turnover, frequent product changes within a FMCG company and the desire to forecast the promotional product demand volume this section describes forecasting methods that might be applicable.

Before making the connection between forecasting methods and the FMCG characteristics a general overview of the different forecasting methods is given, see Figure 15. A common distinction for forecasting models is subjective or objective. At which a subjective forecasting methods is made qualitatively based on human judgement, objective forecasting methods are made quantitatively based on data analysis (Fildes, Goodwin, Lawrence, & Nikolopoulos, 2009; McCarthy, Tsinopoulos, Allen, & Rose-Anderssen, 2006).

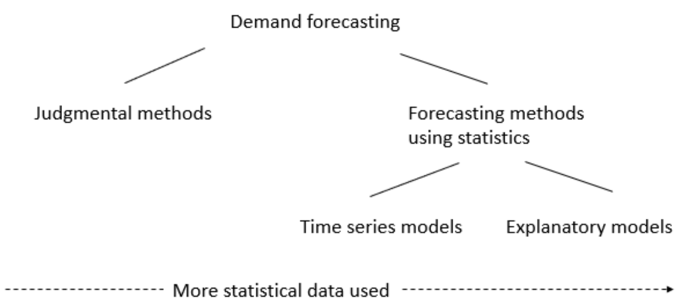


Figure 14: Distinction between demand forecasting models based on use of statistical data. Makridakis (1979); van den Heuvel (2009)

Both subjective and objective forecasting methods will be discussed in this section. Alongside, the combination between subjective and objective methods as well as more sophisticated models are described briefly.

3.2.1.1 Subjective forecasting methods

Judgmental forecasts are made by individual or a group, based on knowledge and experience of the situation to forecast. This research is about forecasting the demand volume of product promotions. Fildes et al. (2009) indicated that the main drivers behind adjusting forecasts are the judgmental adjustments of statistical forecasts for promotional and advertising activities. Commonly used at the company, the "last like" rule is an example of judgmental forecasting. This means that they ordered the same quantity of products that was sold during a similar promotion in the past (Cooper et al., 1999). In practice, this "last like" rule is performed by comparing the volume of prior similar promotions in multiple Excel files to determine the volume forecast of the promotion, based on human judgment.

In line with the findings in Section 2.6, one of the problems causing an inaccurate forecast is organizational influence. This is in line with Fildes et al. (2009) which states that in many companies, senior managers adjust forecasts without consulting the forecasters, possibly for political reasons. However, multiple research has been carried out into the effectiveness of these adjustments and suggests that they can improve accuracy when forecasters have important information about the products they are forecasting that is not available to the statistical method (Syntetos, Babai, Dallery, & Teunter, 2009). At the same time, adjustments made in absence of important information may result that the forecaster might read false patterns in the data and these adjustments are likely to affect the accuracy (O'Connor & Webby, 1996; Armstrong, 2001; Lawrence, Goodwin, O'Connor & Önkal, 2006; Fildes et al. 2009). Despite this criticism human influence have been shown positive as well (Lawrence et al. 2006).

Qualitative techniques are primarily used when data availability is low. For example, when a product is first introduced into the market. Then, human judgement is used to turn qualitative information into a quantitative forecast (Chambers, Mullick, & Smith, 1971). According to Yelland (2010), forecasting for new products is often performed by selecting appropriate 'predecessor' or 'similar' products to the one to be forecasted. In any case, a product forecast based solely on similar products will be highly uncertain and should be compared to and updated based on the initial sales of the new product (Syntetos et al., 2016)

3.2.1.2 Objective forecasting methods

Objective forecasting methods are statistical methods that are used to deliver a quantitative forecast about the future which uses numerical measures and prior data in order to forecast future promotions. These techniques are based on mathematical models and are predominately objective. Three types of objective forecasting methods are described below:

Time series model

Time-series models examine historical data patterns and forecast the future based on these underlying data patterns. Note that the extrapolation of historical demand data patterns into the future is done in the belief that historical demand data represents the future demand data. The most common time-series models are; simple and weighted moving average, trend/pattern projection, and simple mean and exponential smoothing. A time series is a set of values each observed at a specific time either recorded continuously or with fixed intervals. (Brockwell & Davis, (2006); Thomé et. al (2018))

Time series modelling involves analysis of a dynamic system based on input and output data series, which are related to a function. Time series techniques can essentially be divided into two sets of methods: univariate and multivariate (Thome et al, 2018). Univariate is the analysis of a single variable, while multivariate analysis examines two or more variables. Most multivariate analysis involve a dependent variable and multiple independent variables. These techniques are appropriate when the aim is to identify general patterns or trend, without regard to the factors influence the forecasted variable (Armstrong, 2012).

Univariate forecasting methods are based on time series techniques that analyze past sales history in order to extract a demand pattern that is then project into the future (Makridakis, Wheelwright, & Hyndman, 1998). This kind of forecasting techniques are well-suited for companies that handle a large amount of SKUs and where forecasts are desired to be made semi automatically. However, these methods are not able to include additional potentially relevant information, which is key to forecasting promotions. Product promotions aim to change customers demand, since the demand pattern of the time series is thereby also impacted, the time series model is inadequate for forecasting promotional demand.

Explanatory model

Explanatory models are causal forecasting models try to identify relations that were relevant in the past and then apply them in the future. Causal models assume that the variable that is being forecasted, the dependent variable, has a causal relation with the other independent variables. The forecasts of these models are based on this causalization. Linear regression is one of the simplest forms of an causal model. A regression line forecasts the dependent variables based on the selected number of independent variable.

Quantitative forecasting methods are relatively easy to predict based on their underlying information. Without many complications any person can easily forecast based on available data. However, the main disadvantage of this forecasting method is its dependence on data. An error in the available data can lead to wrong forecasting. Therefore, these methods can only be used if proper data is available.

One way to deal with the problem of promotional forecasts involves the use of multivariate statistical models that include information on past promotions for building causal models based on multiple linear regressions whose external inputs correspond to the promotion features (discount price, display, advertising, etc.) (Trapero, Pedregal, Fildes, & Kourentzes, 2013; Cooper et al., 1999). Armstrong (2012) stated that explanatory model is the preferred forecasting method if information about those promotion features is available.

Sophisticated models

Besides time series models and explanatory models there are more sophisticated forecasting models, like support vector machines, neural networks for producing business forecasts. Neural networks have the advantage that can approximate nonlinear functions (Chen, 2011). Whereas the classical methods used for time-series assume that there is a linear relationship between the inputs and output. Due to the proven track record in practice of the time series extrapolative methods they remain very attractive. As well as their relative performance compared to the more complex methods. Makradakis, 2018 concludes that the artificial intelligence methods do not outperform the classical statistical forecasting models.

Furthermore, time series and explanatory methods are quite intuitive, which makes them easy to define and use, and enhances their acceptance by the end-users (Dietvorst, Simmons, & Massey, 2014; Alvarado-Valencia, Barrero, Onkal, & Dennerlein, 2016). Complex methods, such as many machine learning algorithms, often appear as black boxes, and provide limited or no insights into how these forecasts are produced and which data elements are important. These attributes of forecasting are often critical for users (Sagaert et al. 2018). This is in line with the practical requirement mentioned in Section 1.4 prescribes an easy and interpretable model, building on existing tools.

Many quantitative forecasting methods found in literature are performed on time-series models and comprise of forecasting total demand rather than focusing on promotional demand forecast. Therefore the findings and results of the models used in the M4-competitions (Makradakis et al., 2018) cannot be applied directly to this research. However, they might provide some insights and guidance for improving the current forecasting model.

3.2.1.3 Combination of judgmental and statistical forecast

According to Ali, Sayin, Van Woensel, & Fransoo (2009), in some case statistical forecasting methods may face difficulties where judgmental adjustments can play an important role. First, statistical methods may tend to adjust slowly when changes in actual demand occur, depending on the relative weight on historical data compared to the most recent data. Second, when historical data is not or limited available. Third, events that occur at a certain point in time can strongly influence actual demand in that period in time. Whether this event should be seen as outlier, or taken into account as change in demand, will influence future forecasts.

Combining statistical and judgmental forecasts outperforms a single method in terms of forecast accuracy since ‘models and managers have complementary skills’ (Blattberg & Hoch, 1990). Trapero et al. (2013) and Syntetos et al. (2016) compliment this statement but are critical on the extent of these judgmental adjustments since the research on these combined forecasting methods is limited. Due to the complexity of promotions and since promotions are affected by many aspects, combining both statistical and judgmental forecasting methods may be desired depending on the forecast horizon and the demand history (Figure 16)

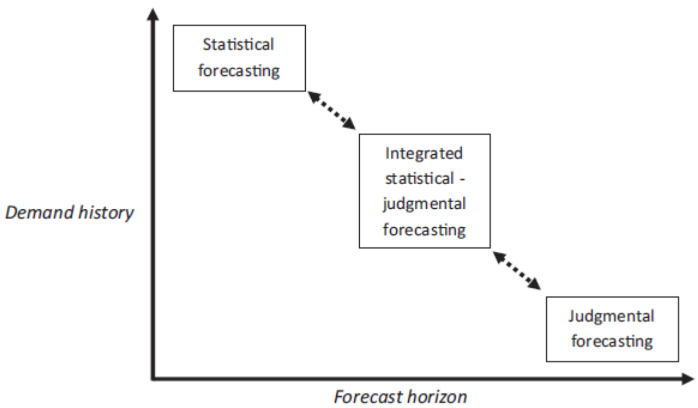


Figure 15: Forecasting techniques in relation to the demand history and forecast horizon – (Syntetos et al., 2016)

3.2.2 Selecting the forecasting methods for product promotions

As already stated in Section 3.2.1.2 time series values are observed at a specific time recorded either continuously or with a fixed interval. As product promotions occur infrequent, and on event basis, according to Van Donselaar et al. (2006) time series seem not appropriate for promotional demand forecasting.

Furthermore, when focusing on historical time series consisting of both base and promotion demand this might not be a reliable representation of future demand because those promotion demand in the past might not be the same in the future demand forecasting as promotions occur infrequent and the demand volume depends on promotion specific variables. These promotion specific variables are not explicitly taken into account as input in time series models.

Explanatory models forecast a dependent variable based on one or more independent variables that have a causal relationship with the (dependent) variable forecasted. This focus on the causal relationship between the variables is desired for promotional demand forecasting since the demand is expected to be dependent on the parameters, which describe the promotion characteristics, that can be modelled by including it as variables. These variables have to be fitted in a model. The most widely used method found in literature is a multiple linear regression analysis (Van Loo, 2006, Van den Heuvel, 2009, De Schrijver, 2009, Cooper et al, 1999, Wittink et al, 1988). In such an analysis multiple independent variables predict one dependent variable. Interaction effects between independent variables can be incorporated when the form of the interacting variables is continuous. Furthermore, the (in)dependent variables can be included in their linear and logarithmic form as long as their form is metric. Traditional methods are insufficient in processing large volumes of data. Therefore more sophisticated modeling and algorithms have been developed to address this problem. Some companies have started to use machine learning methods to utilize the massive volumes of data they already gathered in order to better understand the connections and causality. Machine learning can make it possible to recognize shared characteristics of promotions and identify their effect on the demand sales. To do so, machine learning models use simple versions of (non)linear functions to model complex nonlinear situations.

Van Loo (2006) analyzed the four most important forecasting techniques on criteria that are also applicable to the forecasting environment of Unilever (see Table 1) and rated the relative applicability. A score of 4 indicates that the forecasting method performs relatively the best compared to the other 3 methods and a score of 1 indicates it performs the worst relatively. However, since the development of the more sophisticated methods like Neural Networks it is questionable if this research by van Loo (2006) is still valid. Nevertheless, recent results from the M4-competition by Makridakis, Spiliotis, & Assimakopoulus (2018) show that the artificial intelligence methods did not outperform the classical statistical forecasting models. That is why the research by van Loo (2006) is still assumed valid because the single equation model might still be suitable. Especially since important criteria for Unilever are the ease of use, ease of interpretation and it is important that it is not seen as a 'black box' by the forecasting employees. Single equation models can be further split into single and multiple linear regression models. Since single linear regression models only include on independent variable and in this research the forecasting the volume demand of product promotions depends on more than one independent variable, multiple regression is chosen to be the most suitable forecasting method. This conclusion is in line with prior research of van der Poel (2010).

Table 1: Performance forecasting techniques on several criteria (Van Loo, 2006)

Criteria	Single-equation (single and multiple linear regression)	Multiple-equation	Econometric models	Artificial Neural Networks
Accuracy	1	2	3	4
Costs	4	3	2	1
Complexity	4	3	2	1
Data need	4	3	2	1
Ease of interpretation	4	3	2	1
Ease of use	3	2	1	4
Total	20	16	12	12

3.2.3 Promotion forecasting models for FMCG:

The fact that forecasting events, like promotions, is important in the FMCG sector is evident given the number of comparable researches conducted on this topic, which will be discussed in this section. The results of these comparable researches are shown in Appendix K. In order to gain insights in forecasting models this section provides a brief overview of the forecasting models that already exist. Looking at the existing promotion forecasting models it can be concluded that they are all explanatory models. Cooper et al. (1999) developed a store-specific forecasting model from a retailer's perspective, called PromoCast™. Based on regression, with the natural logarithm of units sold in specific store as dependent variable. It uses many historical averages from a database at SKU level to build a regression model with 67 independent. The model is calibrated for a dataset over 150k SKU's with non-perishable food products. New products are not taken into account. It is hard to reflect total sales during a promotion event. Cooper et al. (1999) argues that the historical averages matching the planned advertising and display conditions provide a benchmark superior to the widely used 'base times lift' method. In comparison with the research at Unilever the focus is the other way around focusing on the parameter conditions providing the lift factor and compare the total demand volume with comparable historical promotions with similar underlying parameter conditions rather than looking at historical moving averages without knowledge of the underlying parameter conditions. The SCAN*PRO is an explanatory model forecasting store and brand specific sales developed to analyze the promotion effects of actual sales via scanning data on store level (Leeflang, van Heerde, & Wittink, 2002).

Both van Loo (2006) and van den Heuvel (2009) used a supply chain wide factor to determine the lift during promotions, the latter uses an average baseline of the past five weeks of baseline demand multiplied by a lift factor determined from several promotion variables. Divakar et al. (2005) developed a decision support system including a sales forecasting tool useful for both base and promotions forecasting for a non-perishable food manufacturer, CHAN4CAST. The forecasting model is built specifically for multiple channels, regions and also major customers accounts are included. The model consists of multiple regression equations for each promotion variables that form the input for the final forecast equation of promotional demand. This model is built from a manufacturer's perspective and focuses on retailer order, where by far most research focused on forecasting shopper demand. Alongside, Divakar et al. (2005) developed a web-based decision support system (DSS) useful for forecasting both base and promotional demand. This DSS includes the effect of promotional variables, seasonality, trend, past sales and significant holidays or new product introductions. The usage of the DSS would be significantly enhanced if it allows users to easily import information. This is consistent with Morris and Marshall (2004), who found that timeframe, feedback signal, and feedback duration are important factors that represent a user's perception of control.

Kock (2012) differs from the other researches as it compares the forecasting ability of multiple linear regression model with more sophisticated data mining forecasting techniques such as neural networks. The lift factors for promotions for non-perishable goods are forecasted from a manufacturer perspective. Kock (2012) shows that the data mining techniques seem performing closely to the linear regression model. This finding endorses the conclusion of Makridakis et al. (2018) – M4 competition, that more complex or sophisticated models are not necessarily more accurate.

Related to the motivation of Unilever for this research to aim for improving the forecast for product promotions Divakar et al. (2005) found similar motivation statements why the forecasting methods used did not meet the senior management expectations:

- Multiple forecasts were generated by the different users such as sales, finance, brand management and the strategic planning based on different methods lacked transparency;

- The forecasts were often inaccurate;
- The forecasts changed frequently for no explicit reason;
- There were no diagnostics or any accountability by the personnel when actual sales deviated from the forecast sales.

Derks (2015) indicates that getting users to accept and internalize the volume forecast of product promotions has the highest improvement potential. In that research mostly judgmental forecasting method is applied to estimate future promotion demand and estimates are based on historical data. Furthermore, no common way of working is used by the forecasting employees in the case of Derks (2015). This is also the case at this research. Both Divakar et. al (2005) and Derks (2015) suggest that when the data is split per specific product groups the models could be probably better estimated. This outcome is supported by Raghubir, Inman, & Grande (2004) that argues the importance to develop separate sales forecasts by channel, since consumer behavior and sales of consumer packaged goods at different channels are driven by different factors. When there is a gap between the actual volume and forecast volume, they need diagnostic information on the drivers of the gap. These diagnostics are possible only if forecasts are available by product category, channel, region and major customer chain or account within each channel, so that managers can drill down and analyze the diagnostic issues in depth. CHAN4CAST by Divakar et al. (2005) not only extends the literature on forecasting and marketing mix effects for the grocery channel but also show why other channels are quantitatively as important as the grocery channel to overall forecast accuracy. Despite the fact that these other channels do not have the same quality of data as the grocery channel, they were able to develop robust models for these channels. Their approach is generalizable because the challenges involved in the forecasting environment in their research, soft drink sales, are similar to those faced in forecasting the sales of other packaged goods.

Trade-offs include one between the use of aggregate versus disaggregate data and the trade-off between model relevance and sophistication, one between model simplicity and completeness (Van Heerde, Gupta, & Wittink, 2002). Efforts to build sophisticated models have to be balanced against practical considerations such as timely completions of the forecasts and automation of the analysis (Van Heerde et al., 2002). All researchers used for this comparison agree on that model relevance is primary, where model sophistication is secondary. Also, these researchers state that that the (ability for) diagnostics is at least as important as the forecast accuracy.

3.3 Literature used for analysis

This section describes the literature used for analysis.

3.3.1 Dependent and independent variables

Regression analysis is a common approach for modeling the causal relationships between one variable, like promotional demand, and one or more other variables, like discount rate, or price, or other variables. Regression analysis makes a distinction between the variable that is being predicted, the so called dependent variable, and the variables used to predict that dependent variable, the so called independent variables.

Poor selection of variables results in low model fit and thus in an inaccurate forecasting model for the company. In this research the focus is on forecasting the demand volume of promotions. This demand volume will be forecasted as dependent variable and is calculated by multiplying the baseline demand times a lift factor. Similar calculation is performed in Cooper (1999), van Loo (2006), van der Poel (2010) and Derks (2015). By taking the lift factor as dependent variable the promotional demand volume is standardized against the baseline demand volume. As argued in Section 3.2.2 this research uses multiple linear regression, where one dependent variable is predicted with multiple independent variables. As dependent variable the lift factor of a promotion is forecasted. This lift factor represents the promotional demand volume divided by the baseline demand volume of a SKU. Supervised learning is applied in order to distinct the different independent variables. This research will also test the effect of these different independent variables on the dependent variable, the lift factor.

According to Field (2009), a great deal of care should be taken in selecting independent variables for a model because the independent variables included and the way that they are entered into the model can have great impact. Ideally, these independent variables should be selected based on past research (Field, 2009). When the independent variable has no or very little effect on the calculated dependent variable it might be better to leave it out of consideration. However, uncorrelated independent variables are rare, making the method of selecting the independent variables crucial.

Adding more independent variables to a multiple regressions procedure does not mean the regression will be ‘better’ or offer better forecasts; in fact it can make things worse. This is called overfitting. When overfitting an analysis corresponds too exact to a particular set of data, and may therefore fail to forecast future observations accurately.

The addition of more independent variables creates more relationships among them. So not only are the independent variables potentially related to the dependent variables, they are also potentially related to each other. When this happens, it is called multicollinearity. Multicollinearity occurs if there is no perfect (linear) relationship between two or more independent variables (Field, 2009).

Ideally, all independent variables to be correlated with the dependent variable but not with each other. Because of multicollinearity and overfitting a fair amount of work should be done as preparation for forecasting.

The multiple linear regression is argued to be the most suitable forecasting method for this research, according to Kock (2012) although the results of linear regression are easy to explain to layman, and can easily be implemented into a tool, it needs lots of (dummy) variables to take all information into account, therefore potentially creating multicollinearity. What dummy variables are is will be described in Section 3.3.3 and in Section 4.3 when the promotion variables are analyzed.

3.3.2 Previous research on independent variables

This section describes findings from previous comparable research on determining the independent variables to gain insights into which variables to incorporate in the model.

Previous research by van der Poel (2010) indicates the relative importance of the independent variables, used by the marketers, that have an effect on promotional sales. Van der Poel (2010) implicates that a promotion where the consumer gets a free product or premiate has the lowest promotional demand, although the success of such promotion really depends on the type of free product or premiate. This justifies the decision (section 1.6) to first focus on the promotion types with the highest promotion demand, the regular promotions like single price off or buy two get one free. Ramanathan & Myldermans (2010) state that a promotion presented on a second placement (display) maximizes attention of potential buyers. This confirms that items on second placement increase the lift factor, which is in line with van der Poel (2010), and Peters (2012) that both state this effect to be extremely high, whereas Derks (2015) found it to be an important driver but the effect to be medium high. According to van der Poel (2010) Second Placement or display of a promotion in a retailer store is far more important than folder- and TV advertisement. Both van der Poel (2010) and Peters (2012) argue that the number of SKU’s in a promotion negatively affect the lift factor of that promotion.

Limited information is available about the effect of the increase in number of shops where the product is sold in promotion on the total promotional demand sales and lift factor. These insights point us in a direction for the analysis to check whether these conclusions can also be validated in this research.

3.3.3 Dummy variables and minimum sample size

Regression analysis treats all independent variables as numerical. Numerical variables are discrete and interval variables that are directly comparable. Often, however, parameters that include attribute data are included into the regression model. In order to convert this attribute data into numerical data that can be processed within the regression model dummy variables are created, see formula 3-1 for an example. A dummy variable is an artificial variable created to represent a parameter with multiple attribute levels. Dummy variables represent these group of attributes using only zeros and ones. The number of dummy variables created for each nominal variable is equal to the number of attributes of that nominal variable subtracted by 1. For the understandability all attributes will be given a dummy variable in the next example. However, in the model the rule of subtracting 1 from the number of attributes per nominal variable will be applied, since it will decrease the required number of data. This will be explained in the next section. Dummy variables need to be created for the variables: promotion mechanism, cluster, retailer, and promotion type.

To illustrate a dummy variable, for example, looking at data of the largest retailers (Retailer A, Retailer B, Retailer C, Retailer D), we can define these new variables:

- $R_{1,t} = 1$ when *t* represents data of 'Retailer A' and zero otherwise;
- $R_{2,t} = 1$ when *t* represents data of 'Retailer B' and zero otherwise;
- $R_{3,t} = 1$ when *t* represents data of 'Retailer C' and zero otherwise;
- $R_{4,t} = 1$ when *t* represents data of 'Retailer D' and zero otherwise.

Then the regression model is:

$$y_t = a + bt + c_1 R_{1,t} + c_2 R_{2,t} + c_3 R_{3,t} + c_4 R_{4,t} + e_t$$ **Formula 3-1**

As can be seen from this regression formula, dummy variables will impact the number of variables of the regression model. What impact the number of variables used in the regression model will make on the data requirement is described in the next section.

3.3.3.1 Minimum sample size

According to the method by Hyndman & Kostenko (2007), if r is the number of variables, then r parameters are required. Additionally, a regression also has another parameter for the time trend. So the number of parameters of this example regression model is r+1. It is always necessary to have more observations than parameters (Hyndman & Kostenko, 2007). Therefore, the theoretical minimum number of sample observations is r+2. Thus, in this example the minimum number of observations for retailer data is 6. But this number will be sufficient only when there is almost no randomness. In this practical case, substantially more data are required. As the sample size (n) increases, the prediction interval decreases at a rate in the order of the square root of n (Hyndman & Kostenko, 2007).

According to Field (2005), the expected R for random data (Formula 3-2) describes the amount of influence random data has on the variance of the model and it can be calculated by this formula:

$$expected\ R\ for\ random\ data = \ k/(N-1)$$ **Formula 3-2**

Whereby, *k* equals the number of independent variables and *N* equals the total sample size. The goal is to get a small value (<0.1) for the expect R for random data so the random data would have very little impact on the variance the model.

Green (1991) states two rules of thumb (Formula 3-3 and 3-4), one to test the overall fit of the regression model (i.e. test the R^2), and the second to test the individual independent variables within the model (i.e. test b-values within the model).

$$Overall\ model\ = 50 + 8k \qquad \textbf{Formula 3-3}$$
$$Individual\ independent\ variables\ = 104 + k \qquad \textbf{Formula 3-4}$$

Whereby, *k* is the number of independent variables. In this case we want to calculate both the overall fit of the regression model and test the individual independent variables. Therefore, we should take the largest value of these two rules.

3.3.4 Forecasting performance measure

According to Chopra and Meindl (2001), users can use forecasting error analysis to determine whether the current forecasting method predicts the demand accurately. For example, if a forecasting method consistently results in a positive error, the user can assume that the forecasting method is over forecasting the demand and can take appropriate corrective action. Chopra and Meindl (2001) state that measuring forecast errors improves forecast accuracy. But, simply measuring forecast errors on a general level does not provide enough information for setting targets for forecast accuracy and does not allow for finding the areas to develop in this forecasting demand process (Mentzer & Moon, 2005). Therefore, we discuss the several forecasting performance measures and their contribution to overcome this problem.

The current performance measure (see formula 2-1) used by Unilever can take negative values. This occurs when the forecast volume is significant higher than the actual volume. For example, if the forecast volume is 1000 and the actual volume is 50, the accuracy according to formula 2-1 is: $Forecast\ Accuracy\ (Unilever) = 1 - \left(\frac{ABS(1000-50)}{50}\right) * 100\% = -18\%$. These negative accuracy values interfere with interpretation and understanding of the promotion data. This offers an opportunity for improvement by comparing other performance measures.

When looking at a dataset with different variables the error measures should be adjusted for the scale in the data. This is a reason why the root mean square error (RMSE) is no longer the most widely used error measure, since it is not scaled to the data. The mean absolute percentage error (MAPE) is designed to overcome this problem and is therefore the most commonly used accuracy measure (Fildes & Goodwin, 2007). However, no error measure will give a complete picture of the actual accuracy. Therefore, Fildes & Goodwin (2007) advice to use multiple measures to assess different aspects of the performance. Although, Fildes & Goodwin (2007) state in their research on forecast error measures that only 30,2% of their respondents used multiple forecasting measures.

The formula for the mean absolute percentage error (MAPE) where *n* refers to the number of SKU's with a promotions or to the number of promotions, depending on the aggregation level.

$$MAPE\ = \frac{1}{n}\sum_{1}^{n}\frac{(ABS(Actual\ Sales - Forecast\ demand\ quantity)}{Actual\ Sales} * 100\% \qquad \textbf{Formula 3-5}$$

Disadvantage of percentage based errors is that they are infinite or undefined if the actual sales is zero and it will have extreme values when the actual sales is close to zero. Therefore, when the product demand is intermittent, the product demand experiences several periods of zero demand and sometimes the demand is high, Hyndman (2014) recommend to use the mean absolute scaled error (MASE), due to its robustness and versatility. However, this measure suffers the same problem as the mean absolute error (MAE) that in an intermittent demand environment, the model will find a result where a zero forecast may prove to be the 'best'. This understates the argument by Fildes & Goodwin (2007) that multiple forecasting measures should be used to capture the different situations within a dataset. A performance measure that fits the requirement of scaled data and has the benefits of the MAPE is the scaled mean absolute percentage error (sMAPE). Disadvantage of using the sMAPE for this research is that this measure is more difficult to understand and more difficult to interpret in order to use the measure in the decision making process (Hyndman, 2014). Hence, the performance measure MAPE will be used for this research.

In addition to this, the measure R-squared (see Formula 3-6) is applied to determine the forecasting ability of the model. According to Field (2005), R^2 describes the proportion of variance of the dependent variable explained by the regression model. R^2 is calculated using the following formulas:

$$R^2 = \frac{Sum\ of\ Squares\ Residuals}{Sum\ of\ Squares\ Total} = 1 - \frac{Sum\ of\ Squares\ Error}{Sum\ of\ Squares\ Total}$$ **Formula 3-6**

Whereby,

$$Sum\ of\ Squares\ Residuals = \sum\left(y' - \overline{y}'\right)^2$$ **Formula 3-7**

$$Sum\ of\ Squares\ Error = \sum(y - y')^2$$ **Formula 3-8**

$$Sum\ of\ Squares\ Total = \sum(y - \overline{y})^2$$ **Formula 3-9**

Where *y* denotes the actual quantity, *y'* is the forecasted quantity by the model, and $\overline{y}$ is mean of the actual quantities Y. A 'perfect' regression model would result in SSE (Formula 3-8) is zero, and R^2 is 1. If the regression model is really bad, SSR is equal to SST (Formula 3-7, 3-9), reflecting that no variance is explained by the regression, and R^2 is zero. Whereas R^2 tells us how much of the variance of Y is accounted for by the regression model from our sample, the adjusted value tells us how much variance in Y would be accounted for if the model had been derived from the population from which the sample was taken (Field, 2005).

Nevertheless, the *adjusted* R^2 (see Formula 3-10) has been criticized because when looking at a sample it does not imply how well a regression model would forecast a different set of data within the population. According to Field (2005) the formula of the *Adjusted* R^2 is:

$$\text{Adjusted } R^2 = 1 - \left[\left(\frac{n-1}{n-k-1}\right)\right](1 - R^2)$$ **Formula 3-10**

Whereby, n is the sample size, k is the number of independent variables in the model, and R^2 refers to formula 3-6. In this case, when we add more independent variables, increase k, and there is no significant increase in R^2, so no significant increase in the ability of the model to explain the dependent variable, the level of *Adjusted* R^2 will decrease. This drop in *Adjusted* R^2 reflects that when we add more variables to the regression, these variables are not explaining much of the variation of the dependent variable. Thus, this measure can be used to analyze the explanatory power of adding variables to the regression model in order to forecast the dependent variable.

Another measure that explains the performance of variables are Beta coefficients (β). These Beta coefficients (β) indicate the effect size and the direction of a variable within the linear regression model. However, this coefficient needs to be corrected for the different scale of each variable, since it is not a standardized measurement (e.g. Promotion type versus the discount percentage). Therefore, we should use the standardized coefficients. The higher the absolute value of the standardized beta coefficient, the stronger the effect.

3.4 Summary of the literature review

Summarizing the general introduction to promotional forecasting in Section 3.1, promotional forecasting refers to the process that seeks to discover correlations between promotion characteristics and historical demand, in order to arrive at an accurate demand forecasting for future promotion campaigns, which decreases the bullwhip effect (Lee et al. 1997).

Section 3.2 answers the research question A: “What is the best forecasting method to forecast product promotions for a company like Unilever?” by stating that based on the research by Van Loo (2006), we see that multiple linear regression is a good starting point as forecasting method for forecasting the volume of product promotions for a company like Unilever. This statement is confirmed by Derks (2015), Divakar (2005), Leeflang, van Heerde & Wittink (2002), and Van den Heuvel (2009), who all state that model relevance is primary to model sophistication.

In Section 3.2.3 this study is positioned in literature, describing findings of relevant scientific research. In Section 3.3 research question D: “How to measure the performance of the forecast model?” is answered using the research by Fildes & Goodwin (2007), the performance of the forecast should be measured using multiple measures. Substantiated by Hyndman (2014) and Field (2005) the measures MAPE, R-squared, MAE and the forecast error will be used as the performance measures in this study.

This chapter supports and lays the foundation for the analysis. Therefore, Chapter 4 describes the methodology, building on this literature review and the current system analysis as preparation steps for the analysis of which the results are presented in Chapter 5.

4. METHODOLOGY

This chapter describes the methodological approach taken in this study, which is based on the build-measure-learn method by Ries (2011). Based on the foundation laid in the earlier chapters about the current system analysis (Chapter 2) and literature review (Chapter 3), a combination of quantitative and qualitative approaches are used in the data analysis. The research data in this thesis comes from four main sources: the external marketing database Nielsen, the database (VisualFabriq™), semi structured interviews with users, and knowledge by a company expert.

The goal is to improve the result of the equation in Figure 17. The chosen approach to achieve this improvement will be elaborated in Section 4.1 to 4.5.

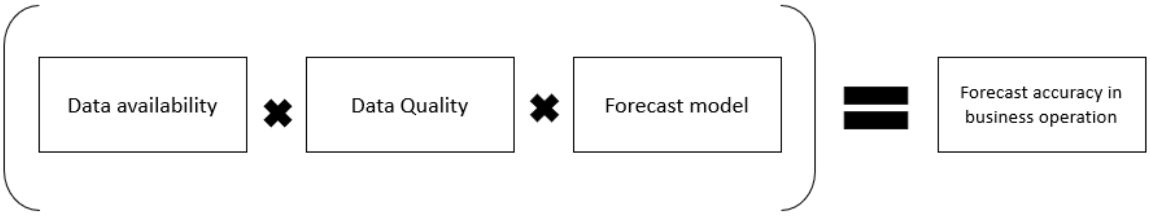


Figure 16: The performances of the different areas of the forecasting method are multiplied, resulting in the forecast accuracy in business practice.

As stated in Section 2.6, the time and the means of this research are limited. Therefore, not all problems can be solved. This requires prioritizing the problem according to the largest improvement potential within the extent of this research. That is why Section 4.1 prioritizes the problem areas found in the root cause analysis (Section 2.6): data availability, data quality and forecasting method. The outcome of this prioritization is used as input for the next section, how we arrive at this outcome is described in detail in Section 4.1. To give a clear overview of the structure of this chapter, the outcome of prioritization, data availability, is already used. Section 4.2 describes how this outcome, data availability, could be improved. Section 4.3 explains the steps that are taken to prepare the data input in order to test the potential improvements. As a consequence, Section 4.4 explains the methods used to distinguish the influence of the input variables and to determine which variables should be selected as input for the applied forecast method described in section 4.5. Section 4.5 compares the applied forecast method with the current forecast method. The relations of the five sections in this chapter is given in Figure 18.

Figure 17: Structure of the methodology

The objective is not to optimize one single area but the get the overall best forecast accuracy of the entire forecasting method, which includes the acceptance and internalization of the forecast by the users. This study aims to find the best outcome of the equation where the performances on the areas data availability, data quality, and forecast model is multiplied, resulting in the forecast accuracy in business operation, this equation is illustrated in Figure 17. The next section will discuss which area of the forecasting method should get priority by focusing on the area which has the largest improvement potential for improving the forecast accuracy in business operation.

4.1 Prioritizing the problems

In order to increase the objectivity of this decision making process, a multiple decision criteria analysis (MCDA) is applied. MCDA refers to a process used to give structure to decision making processes that invoke multiple and different stakeholders, and (in many cases) incomplete information (Zionts, 1979).

The applied method is called Analytic Hierarchy Process (AHP), and was introduced by Saaty (1980). AHP is an effective tool for dealing with complex decision making. By reducing complex decisions to a series of pairwise comparisons, the AHP helps to capture both subjective and objective aspects of the decision. A benefit of using the AHP is the additional technique for checking the consistency of the decision maker's evaluations. Thus, the bias in the decision making process will be reduced.

The AHP considers the set of criteria which is based on research by van Loo (2006), Sanders (2003) and the practical requirements of this research (section 1.6), shown as Level 1 in Figure 16, and a set of alternatives which are retrieved from the root cause analysis (section 2.6), as Level 2 in Figure 19, among which the decision has to be made. An important feature of the AHP is that it in general the best option is not the one which optimizes each single criterion, rather the option which achieves the most preferable trade-off among the different criteria. The purpose of this is to reach the highest improvement potential for the goal, in Figure 19 shown as Level 0, an accepted forecast for product promotions.

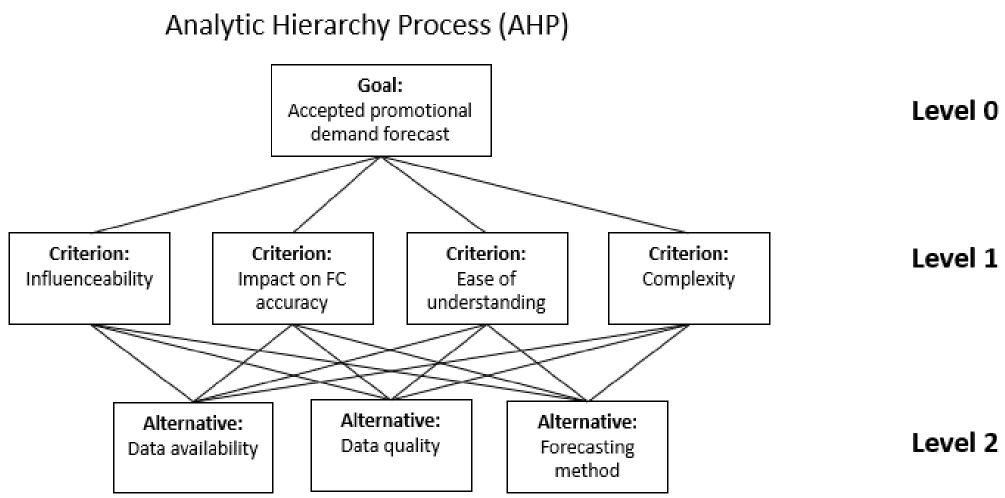


Figure 18: The analytic hierarchy process (AHP)

First, for each criteria a weight is generated based on decision maker's pairwise comparisons of the criteria. For these pairwise comparisons the decision maker uses the values in Table 2. The higher the weight, the more important the particular criterion. Next, for each criterion, scores are given according to pairwise comparisons of the options based on that particular criterion. The decisions on the pairwise comparisons are made with help of the expert within the company, substantiated with practical examples. In the end, these criteria weights and the scores of options on those criteria are combined, resulting in a ranking of the options with respect to the importance, in order to reach the goal.

Table 2: Scores for pairwise comparisons between criteria and alternatives. - Saaty (1980)

Value of a_{ij}	Definition
1	*i* and *j* are equally important
3	*i* is slightly more important than *j*
5	*i* is more important than *j*
7	*i* is strongly more important than *j*
9	*i* is absolutely more important than *j*

Using pairwise comparisons translates the complex decision making process into single evaluations between two criteria or options. Although, these evaluations are simple, the number of pairwise comparisons grows quadratically with the number of criteria and options. This results in a unreasonable workload when evaluation many criteria and options. In this case, see Figure 16, we compare 5 alternatives on 4 criteria. Resulting in 4 * (3/2) = 6 comparisons for building the weight vector of the criteria. And, 4*(5*4/2) = 40 comparisons are needed to build the score matrix on those criteria.

Table 3 represents a $n*n$ pairwise comparisons matrix, where n is the number of criteria. a_{ij} represents the importance of the i^{th} criterion relative to the j^{th} criterion. If $a_{ij} > 1$ then the i^{th} criterion is more important than the j^{th} criterion. Where $a_{ij} < 1$, i is less important than j. It holds that $a_{ij} * a_{ji} = 1$.

The proposed criteria are a combination of criteria found by van Loo (2006), where he studies which forecasting method would be most applicable, and the practical requirements of this research, which state that ease of use and building on existing models is required (see Section 1.6 Scope). Also, Sanders (2003) denotes that his research amongst 240 managers 85,8 percent of them identified the ease of use as most important criteria when implementing software, followed by easily understandable results at a second place with 83,3 percent. A combination of the practical requirements (Section 1.6), Van Loo (2006), and Sanders (2003) result in the following criteria: influenceability, impact of forecast accuracy, ease of understanding, and complexity. These criteria are given a score using pairwise comparisons in order to determine the weight of the criterion. The weight of the criterion reflects the relative importance of that criterion for this research. In this case, influenceability reflects the degree of influence this study could have on the existing organization (e.g. ‘political’ influence, change in way of working). The criteria impact on the forecast accuracy is explained as the potential influence the alternative might have on the performance measures of the forecast, since improving the forecast is the aim of this study. Sanders (2003) explains the criteria ease of understanding by stating, forecast practitioners can better defend their forecasts towards senior management if they understand how they are made. The criteria complexity refers to the degree of factors involved which should be taken into account by both the user and the computer.

The goal of this research is to improve the current forecasting method. As stated earlier in this research the forecast method is likely to improve if the demand forecast for promotions is accepted by the users. Users should get the feeling that the outcome of the forecast method corresponds with the actuals in order to accept it. It might be that then the forecast accuracy will increase, because if users use and internalize the forecast method the quality of input will increase. This is why, the impact on the forecast accuracy is an important criteria. Next, the aim of this research is to improve the existing forecasting environment of the company rather than redesigning the whole forecasting process. Starting the implement a forecasting method from scratch would imply a whole different outline of the research. Also, the influence of this research on the entire company’s forecasting process is limited, so a certain specific area should be picked to focus on. Therefore, the criteria influenceability and complexity are taken into account as important criteria. Next to this, recall section 1.6, the criteria ease of understanding and complexity are also taken into account. The result of these pairwise comparisons of the criteria is given in Table 3.

Table 3: Pairwise comparisons to determine the weights for each criteria

Weighting the criteria		A	B	C	D	Priority Vector
Influenceability	A	1	3	7	9	57,39%
Impact on FC accuracy	B	1/3	1	5	7	29,13%
Ease of understanding	C	1/7	1/5	1	3	9,03%
Complexity	D	1/9	1/7	1/3	1	4,45%
	Sum	1,59	4,34	13,33	20,00	100,00%

The priority vector is obtained by the normalized eigenvector of the matrix, see Appendix B for this calculation and normalized matrix. Table 3 shows that the scores of the options based on the criteria influenceability and impact on forecasting accuracy should be weighted the highest, about 57% and 29% respectively.

Next, each combination of options (Level 2 in Figure 19) is tested using pairwise comparisons with respect to one criteria factor at the time. To start with the criterion with the highest priority vector, influenceability. Subsequently, the criterion impact on forecast accuracy. The same pairwise comparisons could be performed with respect to the factors 'Ease of understanding' and 'Complexity'. However, the weight for these two factors are small (see Table 3, about 9% and 4,5% respectively). Therefore, we assume the effect of leaving these two criteria out from further considerations is negligible. The weights for these two criteria factors is set to zero. The weight factors for 'Influenceability' and 'Impact on FC accuracy' must be adjusted so that the sum will be 100%, see Table 4 (see Appendix B for the calculation).

Table 4: Adjusted weight factor so that the sum will be 100%

	Influenceability	Impact on FC accuracy
Adjusted weighting factor	0,663	0,337

Scoring the options based on the criteria is a subjective process. However, we try to substantiate this decision making by providing the information on which the decisions are made in order to increase the objectivity of the decision making. That is why information about the data availability, data quality, and forecasting method is provided in the next sections.

4.1.1 Data availability

This section briefly describes the data availability of the current forecasting method in order to compare it with the other options based on potential impact on the goal of an accepted demand forecast for promotions. Data availability in this context is considered to be the % of data available per variable and the number of comparable data points, both measured at the moment of promotion confirmation. Measured from a total of the data of all input parameters for each underlying SKU within a promotion.

In the implementation phase of the current forecasting model a lot of push back is given by the users. Users encounter many issues with working with the tool and lack knowledge of the impact of the parameters they have to fill in for a promotion. Next, there is no one way of working and especially if users have an issue they will use multiple workarounds to still be able to deliver their demand forecast. Although it is hard to define variables to the acceptance of users with working with the tool it is possible to define this problem by looking at the degree of correct use of the process flow and input parameters filled in. A first analysis found that most users do not know what the impact is of the lack of information on certain parameters to a promotion. All this results in users not using the tool.

To illustrate, an user complained about the forecasted promotion volume of the promotion 'Soep in Zak, week 44, Retailer A' being twice as low as its own expectation. After a check on the input parameters for the promotion it turned out that the variable '2nd placement' was not filled in. After adding the information on this variable and refreshing the forecast, the forecasted volume became two times higher, in line with the users' expectations. This confirms that, if the information of promotion parameters is not entered into the tool it is not even possible to check whether it is correct or not and will result in the model not being able to forecast the promotion within acceptable accuracy limits. The concept that nonsense, or flawed input data produces nonsense output is described as 'Garbage in, garbage out' (GIGO). Nonetheless, some training materials on the input parameters and the tool are available, although somewhat outdated.

Given this information, data availability is important in order to increase the forecast accuracy. Furthermore, data availability relatively outperforms the other options when it comes to the score of influenceability.

4.1.2 Data quality

This section will discuss a score of the status of data quality in the current process of forecasting product promotions. According to Cheng (2018), data quality consists amongst others of the components: validity, consistency and timeliness. Validity refers to the extent to which data is considered valid and true. Consistency denoted whether various dataset facts match. Lastly, timeliness represent the extent to which data is adequately updated for its purpose.

Because of the fact that users do not properly follow the workflow, the timeliness of the data input is not always correct. This is caused by both the users and the retailers, since information is sometimes shared at the last moment in the process. In order to check the credibility of the data reviewing is included in the workflow. However, the reviewing of promotions is very limited. No or limited sense checks are performed to check whether the forecasted volumes eventually matched with actual demand in that promotion period and to check whether all parameters were correct. If something in a promotion might have changed then this is valuable information when reviewing the promotion to take this information into account when forecasting similar promotions in the future. The interface of the forecasting tool also has some limitations regarding the quality of input data. A user is likely to make a mistake when entering all the parameters of a promotion, but some mistakes could be prevented by the user interface. For example, it is possible to enter a negative demand volume as a forecast for a SKU, which makes no sense in reality. Currently, there is no method or techniques applied in order to check the quality of data input and to ensure its consistency. Data quality highly depends on data availability, since the assessment of the quality of the data is limited by the presence of data. In the previous section is argued why data availability is not perfect, therefore when only looking at the part of data quality the score is corrected for the given level of data availability. So, a comparison of the data quality score can be made between different datasets. Compared to the current situation, an increase in data quality is likely to increase the forecast accuracy.

4.1.3 Forecasting method

The aim of this section is to give a score to the current forecasting method. As a measure for this score the performance of the forecasting method based on the current Supply Chain KPI's forecast accuracy and forecast bias is compared relatively to performance of comparable forecasting models from literature, previous research at the company and with supply chain actuals. Despite the fact that these methods or performance measures might differ in detail and applicability it is known that this comparison is not valid. However, in this stage this comparison will act as a judgement call and point us in the direction whether or not this section has the highest improvement potential and should therefore have to focus of the analysis. **For confidentiality reasons these values are hidden.

*Table 5: Comparison of Forecast Accuracy between three models***

Forecast Accuracy	Foods	Refreshment	Home care	Personal Care	Overall
Van der Poel (2010)	42,3%	n.a.	29,1%	29,1%	31,4%
SC Actuals					
Forecast model (P4)					

In Table 5 the forecast accuracy of the three compared models is shown. Van der Poel (2010) refers to previous research on forecasting promotions at Unilever. SC actuals and the Forecasting model both refer to the current situation up to the first 4 (out of 13) periods of 2019, since the information of that period is at the moment the best representational data. Compared to these other two models the improved forecast model up until the fourth period has the highest score on accuracy based on an aggregated cluster level. It should be stressed that these comparison between the models is done only for a brief validation which problem areas to further investigate first. The SC actuals include the manufacturer and phasing behavior and therefore in the current promotion forecasting environment will be lower than the in-market consumer perspective based on scanning of actual sales.

*Table 6: Comparison of the Forecast Bias between three models***

Forecast Bias	Foods	Refreshment	Home care	Personal Care	Overall
Van der Poel (2010)					
SC Actuals					
Forecast model (P4)					

From Table 6 can be carefully concluded that the SC actuals tend to over forecast and the forecasting model tends to under forecast. Based on more in depth analysis in the current forecasting model it can be shown that the forecast accuracy is significantly higher if all data input is available and reliable, meaning in this situation the right information at the right time. The forecasting method has to deal with the incomplete or incorrect data, described in the previous sections.

At the current stage, the forecast method should perform equal to or better than the current performance to gain trust and acceptance from both users and senior management. The responsibility for the forecast performance is within the Integrated Operations department. However, the responsibility of the quality of the data is not within this department. That is why for this research, the forecasting method is found to be slightly more important than data quality based on the influenceability, but less important than data availability. On the criteria impact on forecast accuracy data availability is said to be more important than the forecast method.

First, these three sections describing the options with respect to influenceability result in the pairwise comparisons of the options shown in Table 7. The detailed calculations are shown in Appendix B.

Table 7: Pairwise comparisons of options, scored with respect to the factor influenceability

With respect to the factor Influenceability		X	Y	Z	Priority Vector
Data availability	**X**	1	7	5	**72,35%**
Data quality	**Y**	1/7	1	1/3	**8,33%**
Forecasting method	**Z**	1/5	3	1	**19,32%**
	Sum	**1,34**	**11,00**	**6,33**	100%

With respect to influenceability, the option data availability is given the highest score (see Table 7). The calculations of these priority factors are given in Appendix B.

Subsequently, the options are compared and scored with respect to the impact on the forecast accuracy, see Table 8.

Table 8: Pairwise comparisons of options, scored with respect to the factor impact on FC accuracy

With respect to Impact on FC accuracy		X	Y	Z	Priority Vector
Data availability	**X**	1	5	7	**72,35%**
Data quality	**Y**	1/5	1	3	**19,32%**
Forecasting method	**Z**	1/7	1/3	1	**8,33%**
	Sum	**1,34**	**6,33**	**11,00**	100%

With respect to the impact on the forecast accuracy, data quality is the most important at this stage, given the current forecasting method (see Table 8). In order to increase the forecast accuracy the data quality should be improved. However, to get an overall accepted forecast, first the data should be available in order to improve the quality of that data. Second, a good forecasting method is required that can cope with the given data availability and quality level. The calculations of these priority factors are given in Appendix B.

Conclusion of the analytic hierarchy process

To conclude, the analytical hierarchy process by Saaty (1980) is applied in order to prioritize problems. Combining the weights of the criteria and the scores of the options with respect to those criteria, respectively Table 4, Table 7 and Table 8, results in the following prioritization of the options (see Table 9).

Table 9: Prioritization of the options based on a combination of the criteria weights and scores of the options with respect to those criteria.

	A	B	Priority factor
Adjusted weight factor	*0,663*	*0,337*	
Data availability	0,480	0,244	72,35%
Data quality	0,055	0,065	12,03%
Forecasting method	0,128	0,028	15,62%

However, given the limited amount of time and influence of this research the focus is first on data availability. And, based on the potential impact of the forecast accuracy data availability is also found to be the alternative with the highest priority factor and therefore should be given priority.

Each pairwise comparisons matrix is checked for consistency in the decision making process. For the readability of this report, these calculations are shown in Appendix B. According to calculations of both the consistency index as the consistency ratio, the decision making in each pairwise comparisons matrix used in this research (Table 4, 7-9) is found to be consistent, see Appendix B. Based on this AHP we can conclude that 'Data availability' is the most preferable option to prioritize on with a score of 72,35%. Followed by 'Forecasting model' with a 15,62% score (Table 9). Therefore, in order to reach

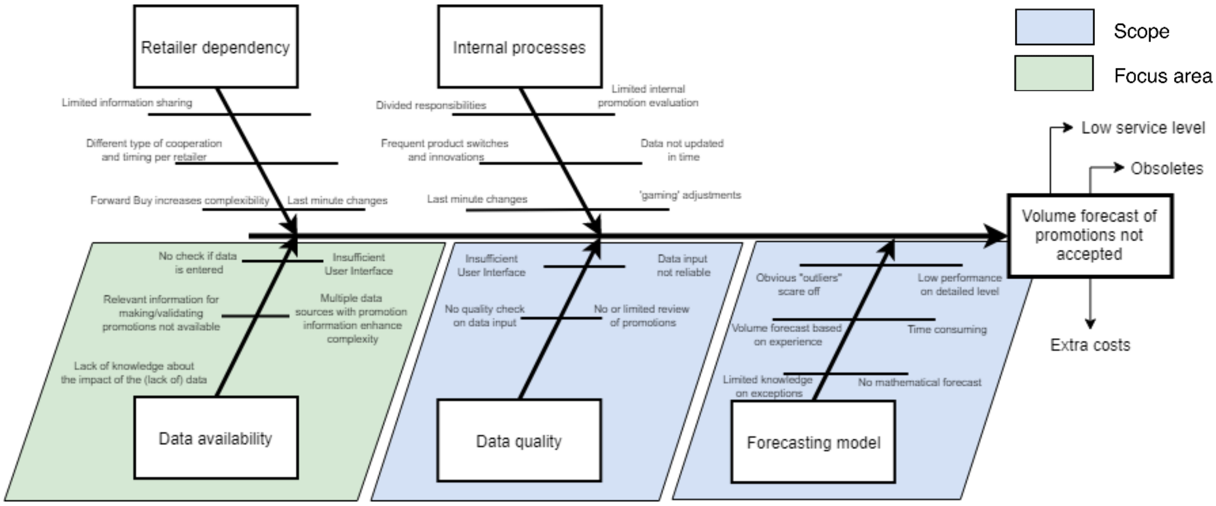


Figure 19: Prioritization of the problem area's

the goal of improving the forecast accuracy and getting the forecast accepted by the users, the priority should be given to 'Data availability', see the green area in Figure 20. Followed by 'Forecasting method' and 'Data Quality', see blue areas in Figure 20. For this reason, in the next section will describe the main drivers of the data availability problems and how we could improve them.

4.2 Improving the data availability

In this section hypothesis are drafted given the prioritization of problem areas in the previous section. The problem area with the highest priority is data availability. In this study, by means of data analytics and semi structured interviews with the users of the forecasting method, both quantitative and qualitative research was conducted to identify the factors that impact the data availability.

Based on these interviews (the details of these interviews are given in Appendix C), the aspects that impact data availability are summarized by the factors; knowledge of the users, the user interface, and the complexity of data sources. According to the users, the user interface causes the greatest frustration. This finding is supported by Fabricant (2013), which states that bad user interface leads to no or less usage. Therefore, the focus is on the user interface. Users are overwhelmed by the amount of information they need to enter into the tool. Besides the amount of input data, the knowledge of which information to enter or the importance of the variables is not always known by the users. In some cases, for example when entering a promotion, the users feel like they have many options or exceptions to consider when filling out the information, which makes the process of entering a promotion more difficult and time consuming. According to Tunikova (2018), this phenomena, the state of feeling overwhelmed by the volume of information to the point at which one feels more confused than knowledgeable about a particular topic, is called information overload. Information overload can increase the difficulty of decision making. Fabricant (2013) substantiates this by stating that successful user experience is often about doing less, not more.

It is desired to analyze the impact of decrease the complexity of the user interface, since this is found to be an answer to the research question F (*How to ensure that the tool is used and internalized?*). Therefore, we try to decrease the information overload for users and increase the user satisfaction with using the user interface. We aim to accomplish this by decreasing the complexity of the model, identify the important variables, and decrease the number of variables. In order to determine which variables should be selected in the simplified model in Section 4.4, first the input data is prepared in the next section.

4.3 Preparation of the input data

Data preprocessing is an often neglected but major step in the data mining process (Garcia & Herrera, 2015). The data collection is usually a process loosely controlled, resulting in out of range values, impossible data combinations, missing values, etc. According to Garcia & Herrera (2015), if there is much irrelevant and redundant information present or un reliable data, then it is hard to gain proper insights and knowledge from this data. This is why we apply the following preprocessing steps before running the analysis: dataset preparation, dealing with noisy data, transformation of the data, and cross validation.

4.3.1 Dataset preparation

The initial dataset retrieved from both internal databases and the external company contained 464 columns with variables, containing the entire model including the 'Ex-factory' and 'Phasing' part, these are removed as described in 1.6, we only have an In-Market perspective. From the remaining 242 columns we only used the relevant information for the forecasting model. Subsequently, we only use the product and promotions dimension relevant for forecasting the product volume of SKU's. This results in a remainder of 72 variables, consisting of 50 numerical variables and 22 object variables. We do not include the variable previous lift factor in the model as it comprises indirect the values of the other variables of a previous promotions, thereby it can affect the significance of the other independent variables as they explain the same variance. Furthermore, the true effect of the variables can be disturbed when keeping the variable that describes the lift factor of previous promotions in the model for analysis (Van Donselaar et al, 2006). This also applies to the variables containing forms of the total units (e.g. total units forecasted by the users). Because of this, only 66 of the remaining 72 variables are included as independent variables. This dataset contains data from 4 different clusters, categories, containing weekly promotions data from 2016 until 2019 week 20, covering 345549 SKU's. The promotions are planned in weeks, therefore we will partition the dataset by week. The promotions differ per cluster and per category. Therefore, we will partition the data per cluster and per category.

For the dataset used for the proposed model we will apply several filtering and preparation steps. Since it is impossible to measure and analyze the entire population, we seek for a subset of data that is manageable and represents a large proportion of the total population. In order to put this focus to the analysis the criteria 'promo pressure' and 'size of retailer' from the section 1.6 Scope are used. Based on both the 'promo pressure' and 'size of retailer', see Table 10 and Figure 21 respectively, we select the data of Retailer A for conducting the analysis because it outperforms other retailers on both promo pressure and retailer size. *Please note, the values and retailers in Table 10 and Figure 21 are anonymized for confidentiality purpose.

*Table 10: Promo pressure for the largest retailers in 2018.**

Promotion Pressure	Foods	Home Care	Personal Care	Refreshments
Retailer A	**26%**	**61%**	**46%**	**29%**
Retailer B	15%	32%	13%	24%
Retailer C	19%	42%	64%	22%
Retailer D	23%	50%	26%	28%

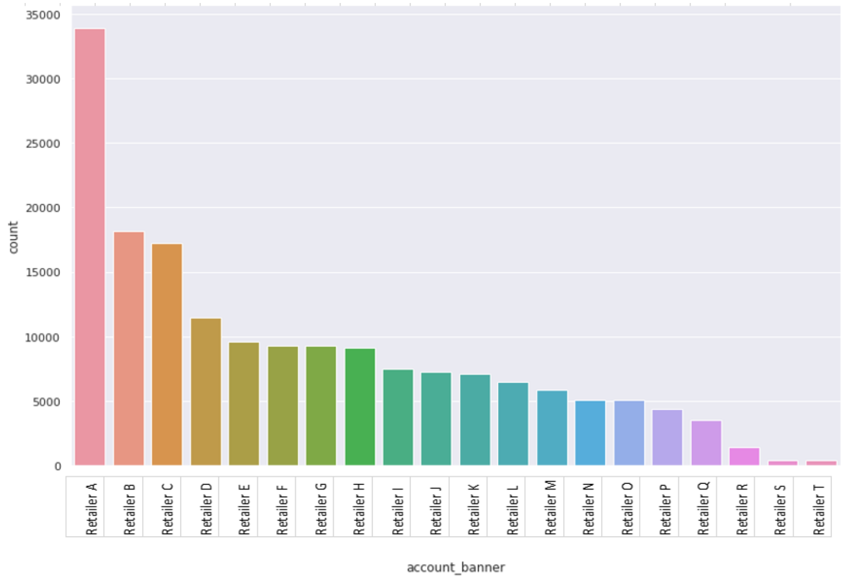


*Figure 20: The amount of quantity sold per retailer, reflecting the size of retailer**

So, for the remainder of this research we only look at the data of Retailer A, unless stated otherwise. That is why only promotions with *account_banner* is *'Retailer A'* will be taken into account, therefore *49772* SKU's from the entire dataset of *345549* SKU's are left. During the review step in the workflow, users can exclude promotions in the past from being used as training data for the model. Users have several reasons for excluding a promotion, e.g. out of stock period or fire in the factory, which might harm the patterns in the data and should therefore be seen as an exception. These exceptions, *4300* SKU's, are removed from the dataset, resulting in a remaining dataset of *45472* SKU's. Subsequently, for this analysis we are only interested in promotions with the scanning data of actual sales, which is described as status 112 or higher. Therefore we only keep SKU's that are in 'promotion_status' higher or equal to status 112, removing *3612* SKU's and this leaves us with *41860* SKU's. As defined in the scope of this research, we only look at regular and category promotions because of their pattern. Therefore, *3129* SKU's are removed, leaving *38731* SKU's. Next, all *1827* SKU's that did not have any quantity in the past promotions are filtered, otherwise this would result in infinite performance scores which affect the overall score. In order to further limit noise to the data some filter steps are added to filter SKU's that do not make sense at first sight. An SKU should have a baseline in order to forecast the volume, since this is a multiplication of the lift factor times zero would result in an error. Lastly, the lift factor and discount percentage are limited to $0.5 < x < 60$ and $0 < y < 70$, respectively. All these filter steps result in a total dataset used for the remainder of this analysis of *35896* SKU's. This dataset contains data from the retailer 'Retailer A' and 4 different product clusters Foods, Refreshment, Home Care, and Personal Care. To illustrate, all these filter steps are shown in Table 11.

Table 11: Applied filter steps to the dataset, resulting in the dataset used for analysis

	Filter	Removed	Left	Units %
0	Starting with Nr Records:	0	345549	NaN
1	Only account banner [redacted]	295777	49772	37.8
2	Filtered Excluded Records	4300	45472	36.6
3	After status 112+ filter	3612	41860	33.2
4	After channel filter	0	41860	33.2
5	After Regular + Segment promotion_type filter	3129	38731	27.5
6	After mechanism_type filter	0	38731	27.5
7	After week 201601+ filter	0	38731	27.5
8	IM Qty should be > 0.0 filter	1827	36904	27.5
9	Baseline should be > 0 filter	242	36662	27.4
10	Less than 5 weeks consumer filter	0	36662	27.4
11	After discount 0 < x < 70 filter	381	36281	27.3
12	Mechanism should be filled filter	0	36281	27.3
13	After lift between 0.5 and 60.0 filter	385	35896	27.2

4.3.2 Dealing with noisy data

Teng (1999) describes three methods of handling noisy data (e.g. missing values). The first method, keeping the noise to prevent the model from overfitting. The second method is to discard the noise beforehand. The third method is to find the noise and try and correct it. The focus of this section is on data availability rather than data quality, therefore from the noisy data we look at the missing values.

Intuitively a missing value (MV) is just a value for attribute that was not introduced or was lost in the recording process. There are various reasons for the existence of missing values, such as manual data entry procedures, equipment errors and incorrect measurements. The presence of such imperfections usually requires a preprocessing stage in which the data is prepared and cleaned. The simplest way of dealing with MVs is to discard the cases that contain MVs. However, this method is practical only when the data contains a relative small number of cases with MVs and when analyzing the complete data will not lead to serious bias. Alternatively, solving the missing value problem by imputing a value (e.g. average). Royston (2004) suggests to create a small number of copies of the data (3 or 5), each with the missing values suitably imputed and after analyzing each dataset independently, the estimates of the parameters of interest are averaged across the number of copies to give a single estimate. Or only use the missing values when they are not observed completely at random. Use ML to impute these missing values. Although ignoring missing values often results in a substantial decrease in the sample size available for analysis, it does have important advantages. In this case, we are dealing with data that is missing at random (MAR). In which the distribution of an example having a MV for an attribute does depend on the observed but not on the unobserved data (Garcia et al., 2015). Therefore, we apply the first method, thus we will keep the noisy data in order to prevent the model from overfitting to the training data.

4.3.3 Transformation of the data

In this steps we check which data needs to be converted from one format into another format, this is called data transformation and it is a fundamental aspect of data processing. The need for transformations is presented in order to satisfy the assumptions under linear regression to have normally distributed error terms and to check the homoscedasticity assumption (Field, 2005). These transformations are applied to transform the distribution of the data variables closer to the normal distribution, which could solve the problem of the rejection of these assumptions. The forecasting method for a simplified model is a linear regression model. Therefore, the data of the dependent variables should be somewhat normally distributed. Looking at the descriptive statistics of the dependent variable *total_units* (see Figure 23) it does not look like the data is normally distributed.

Descriptive Statistics: total_units

Variable	N	N*	Mean	SE Mean	TrMean	StDev	Variance	CoefVar	Minimum	Median
total_units	35896	0	10033	119	6443	22535	507838468	224,61	1	2750

Variable	Maximum	Range	IQR	Skewness	Kurtosis
total_units	620719	620718	8810	7,21	94,64

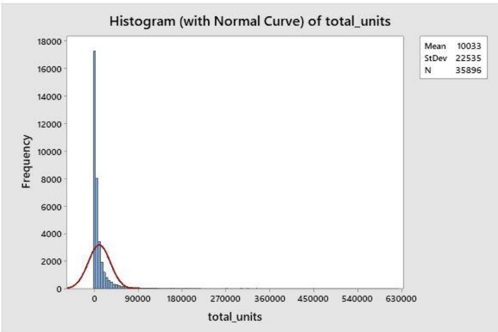


Figure 22: Descriptive statistics of the dependent variable 'total_units'

Descriptive Statistics: LN_totalunits

Statistics

Variable	N	N*	Mean	SE Mean	TrMean	StDev	Variance	CoefVar	Minimum	Median
LN_totalunits	35896	0	7,8041	0,00988	7,8363	1,8720	3,5044	23,99	-0,5738	7,9194

Variable	Maximum	Range	IQR	Skewness	Kurtosis
LN_totalunits	13,3386	13,9124	2,5702	-0,28	-0,31

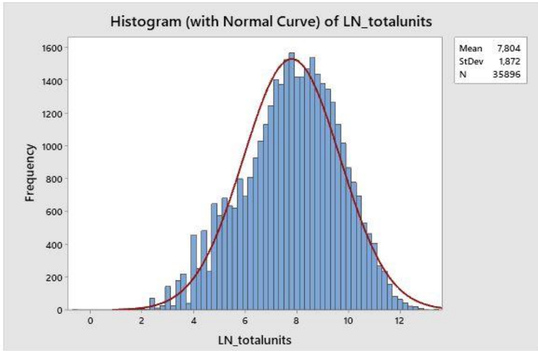


Figure 21: Descriptive statitics of the dependent variable transformed to 'LN_total_units'

In order to fit the data of *total_units* to a normal distribution this variable is transformed using the natural log of the data. The descriptive statistics of this transformed variable *LN_total_units* is given in Figure 22. The transformed variables has a better fit with the normal distribution, therefore the dependent variable *LN_total_units* is used in all analysis.

An overview of the input parameters for promotions at Retailer A is shown in Table 12. Some of these input variables contain explanatory information such as amongst others ‘Cluster’, ‘Brand’ and ‘Category’. Since linear regression models requires only numerical variables, non-numerical variables need transformation as described in section 3.9, using dummy variables. Therefore, according to Field (2005), we can distinct this as numerical and categorical variables and this is denoted in the most right corner of Table 12. These categorical variables are converted into dummy variables using the function LabelEncoder which encodes the variables with a value between 0 and n_classes-1 (Scikit-learn developers, 2018). Therefore, all variables are transformed into numerical variables. For each variable the minimum and the maximum values are given in the columns ‘measurement’, where the price is measured in euro’s, the percentage as %, and one unit refers to one SKU.

Table 12: Overview of all input parameters for product promotions

Variable	Description	Measurement	Scale
2nd placement	Whether or not products of the promotion are placed on a display, 2nd placement in the store.	(0, 1)	Categorical
2nd placement in % of stores	The percentage of the selling stores in which the promotion has a 2nd placement in the store (display).	(1, 99) %	Numerical
Baseline units	Reflects the volume of products sold in a regular, non-promotion, period.	(0.4, 75335)	Numerical
Cluster	Describes the cluster of the products (Foods, Refreshments, Home Care, and Personal Care)	(0, 3)	Categorical
Discount percentage	Describes the percentage of discount the consumer receives for the products	(0, 70) %	Numerical
Discount price	Describes the absolute price discount received by the consumers	(0, 19) €	Numerical
Folder	Describes if the promotion is shown in the folder of the retailer.	(0, 1)	Categorical
Lift factor	Describes the average lift factor of former promotions.	(0.5, 59)	Numerical
Period	Reflects whether or not the promotion is in a 'special' period, like national holiday or (sport)event	(0, 1)	Categorical
Promo Mechanism	Reflect to the mechanism used in the promotion and will be programmed with dummy variables (SPO, X for Y, 1+1, etc.)	(0, 4)	Categorical
Promo type	Describes the type of promotion (Regular, Category, Tailor Made, etc.)	(0, 1)	Categorical
Promo-length	Length of the promotion in weeks	(0, 4)	Numerical
Retailer	Reflects the retailer at which the promotion is held (Retailer A, Retailer B, SU or Retailer C)	(0, 1)	Categorical
Tag	Depict if the promotion is part of a special promotion streak	(0, 1)	Categorical
Total number of products	Describes the total number of products (SKU's) within a promotion	(1, 564)	Numerical

4.3.4 Cross validation

It is common practice to measure the performance of your model with data that is not included in training the model, e.g. to make a split between training data and test data. This method is called cross validation and will provide better insight in the generalization error. This dataset is sparse, therefore we apply a cross validation technique with the least loss of training data. This is we apply the KFold cross validation technique, which is illustrated in Figure 24. A common split of using 80% as training data and 20% as test data is applied.

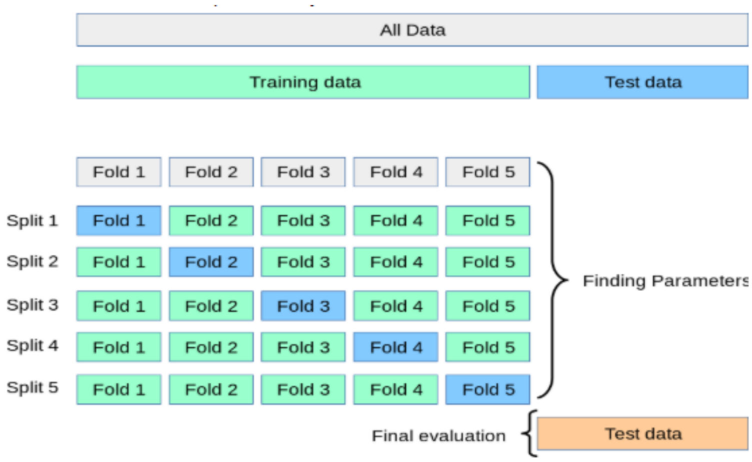


Figure 23: The KFold cross validation technique – Scikit-learn developers (2018)

If the whole data set is used for both build and validate the model generated by a machine learning algorithm, we have no clue about how the model will behave with new, unseen cases. If a model is able to make accurate predictions on unseen data, we say it is able to generalize form the training data to test data (Muller & Guido, 2017). We want to build a model that is able to generalize as accurately as possible. Two problems may arise by using the same data to train and validate the model, their trade-off is visualized in Figure 25.

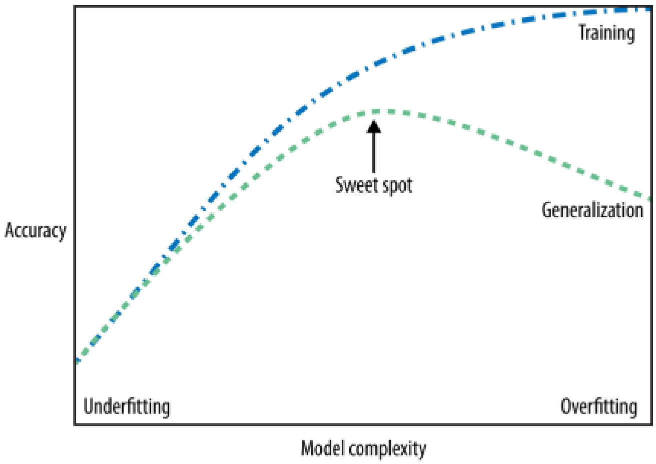


Figure 24: Trade-off between underfitting and overfitting to find the sweet spot for generalization

On one hand, underfitting happens when the model is poorly adjusted to the data, suffering from high error in both training and test (unseen) data. On the other hand, overfitting happens when the model is too tightly adjusted to data offering high precision to known cases but behaving poorly with unseen data. The model learns the detail and noise in the training data to the extent that it negatively impacts its performance on unseen data. Figure 25 shows the trade-off between model complexity and accuracy, where the goal is to find the sweet spot of the generalization. Flexible and complex models have the risk of following the noise too close (overfitting) which results in bad generalization. Inflexible models on the other hand, may have a bias (to the mean) and may miss valuable patterns (Raschka, 2015).

Cross validation is applied, with a split of 80% training data and 20% test data to fit the linear model to the training data in order to test this fit using the test data. The level of fit is determined by the measure R-squared. We analyze the R-squared in three different ways. First, the R-squared of all independent variables (X) is analyzed to check the proportion of the variance in the dependent variable that is predictable using the independent variables. Subsequently, this analyzed is also done for both the training data (X_train) and test data (X_test). If the R-squared of X_test is significantly lower than X_train this might indicate overfitting on the training data. The R-squared of X_test should always be lower than X_train because the proportion of variance explained by the data from independent variables used in X_test should not be higher than those in X_train, since we use a split of 80% X_train and 20% X_test. The result of this split is given in Table 13, where the values corresponds with the number of SKU's used.

Table 13: Number of SKU's used in the dataset, training set, and test set

Dataset (# SKU's)	Training data (# SKU's)	Test data (# SKU's)
35896	28716	7180

The applied linear regression, or ordinary least squares, is the simplest method for regression. Linear regression finds the parameters for the intercept and slope that minimize the mean squared error between predictions and the true regression targets on the training set. The mean squared error is the sum of the squared differences between the predictions and the true values. In the next section will introduce regularization methods that slightly modify the learning algorithm such that the model generalizes better, therefore improves the model's performance on unseen data.

4.4 Variables selection

Because we have prepared the dataset in the previous section, we can construct the features (synonyms for variables or attributes) by collecting them from the dataset. The goal of this section is to select only those independent variables that contribute to forecasting the outcome. A regularization method is used for variable selection.

4.4.1 Regularization methods

Regularization methods for shrinkage and variable selection of linear regressions model are Ridge, LASSO and Elastic-Net. These regularization methods penalize models with many features or prevent the models from selecting too many features (James, Hastie, Witten, & Tibshirani, 2013). These methods are explained in order to select the most appropriate method for this study. The selected method is then applied for variable selection.

The main idea behind both Ridge Regression is to find an function that doesn't fit the training data perfectly well, by introducing a small amount of bias when fitting the function to the test data. In return for the small amount of bias we get a drop in variance. In other words, by starting with a slightly worse fit with the trainings data, ridge regression can provide a better forecasts on unseen data. For multiple variables, then the Ridge Regression Penalty will give a penalty by squaring each parameter, except for the y-intercept. When the sample sizes are relatively small, then Ridge Regression can improve predictions made from new data (i.e. reduce variance) by making the predictions less sensitive to the training data.

The Ridge Regression tries to minimize the sum of the squared residuals + λ * Slope^2.
The parameter λ is found by cross validation and the slope for all variables are taken into account, except for the intercept, y_i. This results in the following formula:

$$Ridge\ Regression = \sum_{i=1}^{N}\left(y_i - \sum_j x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}{\beta_j}^2 \qquad \textbf{Formula 4-1}$$

Ridge regression and LASSO regression are similar, but have some very important differences. The difference between Ridge and LASSO regression is that Ridge regression can only shrink the slope asymptotically close to 0 while Lasso regression can shrink the slope all the way to 0. Therefore, when predicting with multiple variables using Lasso Regression some parameters will go all the way to zero and are therefore eliminated. Since Lasso Regression can exclude useless variables from equations, it is a little better than Ridge Regression at reducing the variance in models that contain a lot of useless variables. Contrary, Ridge Regression tends to perform a little better when most variables are useful.

The LASSO Regression tries to minimize the sum of the squared residuals + λ * |Slope|. Like the Ridge regression, the parameter λ can be any value from 0 to positive infinity and is determined using cross validation and the slope for all variables are taken into account, except for the intercept, y_i. This results in the following formula:

$$LASSO\ Regression = \sum_{i=1}^{N}\left(y_i - \sum_j x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}\left|\beta_j\right| \qquad \textbf{Formula 4-2}$$

A drawback of using Lasso and Ridge regression is when parameters are correlated. Lasso regression tends to pick just one of the correlated terms and eliminates the others. Whereas Ridge regression tends to shrink all of the parameters for the correlated variables together. A solution to this is the use of Elastic-Net Regression, which is a hybrid of both Lasso and Ridge regression. Elastic-Net regression will group and shrink the parameters associated with the correlated variables and leave them in the equation or will remove them all at once. Elastic-Net regression is often applied when we don't know in advance whether some parameters should be important or when the parameters are correlated.

Elastic-Net regression combines the strengths of both Lasso and Ridge regression and is calculated by minimizing the sum of the squared residuals + λ_1*|variable$_1$|+…+|variable$_x$| + λ_2* variable$_1^2$ + … + variable$_x^2$. This results in the following formula:

$$Elastic\ Net\ Regression = \sum_{i=1}^{N}(y_i - \sum_j x_{ij}\beta_j)^2 + \lambda_1 \sum_{j=1}^{p}|\beta_j| + \lambda_2 \sum_{j=1}^{p}{\beta_j}^2$$ **Formula 4-3**

Where λ1 = Lasso regression penalty and λ2 = Ridge regression penalty.

Taking a mathematical approach we define the following options:

- When λ1 and λ2 = 0, then we use the Ordinary Least Squares to estimate the parameters, by minimizing the sum of the squared residuals.
- λ1 > 0 and λ2 = 0, then we get Lasso Regression
- λ1 = 0 and λ2 > 0, then we get Ridge Regression
- λ1 > 0 and λ2 > 0, then we get a hybrid Elastic Net-Regression.

Müller & Guido (2017) state in their research that having some coefficients be exactly zero, due to applying LASSO regression, often makes a model easier to interpret, and can reveal the most important features of the model. Also, this ensures the model is not overly complex and it prevents the model from overfitting. Since the goal of this section is to select the few parameters that significantly influence the forecast we will apply the LASSO regression method when analyzing the model.

Feature selection is performed for the dataset with all variables using a standard linear regression and with the variables after applying LASSO with multiple values for its hyperparameter, alpha. Feature selection is an important step to prevent the model from becoming unnecessarily complex and could result in a poor application due to overfitting. The goal of this feature selection is not to find the optimal value for alpha, but to find an area that gives a direction for further research. An alpha of 1 might indicate underfitting and an alpha of 0 might indicate overfitting (Field, 2005). Therefore, based on personal judgement a set of alpha's in steps of multiplications by 10 are used for this analysis. Thus, the levels of alpha that are used in analysis are {10, 1, 0.1, 0.01, 0.001, 0.001, 0}. An alpha level of 0 implies a regular linear regression with the aim to minimize the ordinary least squares. The list of variables included at each level of alpha is given in Appendix I.

The performance of these cases is measured using the R-squared. Cross validation as described in Section 4.3.4 is applied to these calculation in order to check the deviation in the results. These results are shown in Table 14. The relatively low deviation in the R-squared calculations using a 5 fold cross validation tell us that these calculations are stable, see the most right columns in Table 14.

Table 14: Analysis of the number of features and the R-squared for both training and test data for the models with different levels of LASSO's hyperparameter alpha.

Models:	# features used	R-squared Training (80%)	R-squared Test (20%)	R-squared Cross validation (cv=5)
Lasso(10)	13	0,51442	0,50732	0,51 (+/- 0,06)
Lasso(1)	19	0,65012	0,64588	0,64 (+/- 0,02)
Lasso(0,1)	31	0,69102	0,68696	0,68 (+/- 0,03)
Lasso(0,01)	41	0,71074	0,70685	0,70 (+/- 0,01)
Lasso(0,001)	59	0,72512	0,72199	0,71 (+/- 0,02)
Linear Regression	72	0,72513	0,72197	0,71 (+/- 0,02)

Next, the training data is used to fit the model and therefore train the model on its parameters. Subsequently, this trained model is used to make predictions of the dependent variable. So, depending on the number of features used, see the second columns of Table 14, the model is trained and the R-squared of the prediction of the dependent variable LN_*total_units* is given in the third and fourth column of Table 14. These steps are performed for each different level of alpha with the corresponding number of variables taken into account. The impact data availability of selecting different levels for alpha is discussed in the next section.

4.4.2 Data availability

As mentioned in Section 4.3.2, data availability is considered to be the % of data available and the number of data points for each variable at the moment of promotion confirmation. Measured from a total of the data of all input parameters for each underlying SKU within a promotion. A simple example, if 5 input parameters are measured and the promotion has 5 SKU's, the total number of data that should be filled out is 5*5 = 25. If, at the moment of promotion confirmation, one input parameter is completely empty, the number of actual data is 20. This results in a data availability of 20/25 = 80%.

Measuring the missing values for the different dataset results in Table 17. The number of missing values is divided by the number of data fields used, which is calculated by multiplying the number of SKU's (equal to the number of rows) by the number of variables used (equal to the number of columns).

Table 15: The results of the analysis on missing values for the different levels of alpha.

Applied models:	Linear Regression	Lasso0.0001	Lasso0.001	Lasso0.01	Lasso0.1	Lasso	Lasso10
# of variables used	72	59	55	41	31	19	13
# of missing values	380532	274707	239069	140253	122721	64545	14626
# data fields used	2328150	2081225	1940125	1446275	1093525	670225	458575
% missing values	16%	13%	12%	10%	11%	10%	3%

The findings in Table 17 will be used as a performance measure in the consideration which version of the model to use. Furthermore, the % of missing values is used to compare the different models in terms of potential improvement on data availability in Chapter 5.

4.5 The simplified forecast method

This section describes the applied methodology in order to prepare for the analysis. Since we are trying to forecast a quantity, e.g. the demand volume of product promotions, we are dealing with a regression problem (James et al. 2013). We will perform the regression analysis using machine learning techniques. The forecasted demand volume during promotions are derived by the model, based on the input parameters.

Machine learning is a kind of artificial intelligence where the machine iteratively improves its performances in a given task. Based on the used model the machine will tweak certain parameters (e.g. weights), and tries again. The machine knows when its improving because with each iteration is the machine is scored using a performance measure. This so called learning process is repeated until a stopping criterion is reached or if the model cannot be optimized any further. After training the machine it can be used as a forecasting model by entering parameters for the specific forecasting environment for which it will be applied. The quality of the forecast depends on the input data the model is trained on, the type of model and the performance measures. A simplified version of forecasting method is presented in a flowchart in Appendix F.

In order to build the model we apply the following steps, based on the methodology of multiple machine learning projects (Makridakis et al. 2018; Pathak, 2018):

- Importing the libraries
- Importing the data
- Removing of the unwanted columns
- Encoding the categorical variable
- Splitting the data into train and test set using cross validation
- Fitting the data to the model
- Making the predictions and calculating the performance measures

4.5.1 The simplified model metrics

The applied model uses multiple linear regression analysis based on the features, also called parameters. These features and their beta coefficients are calculated using machine learning. These beta coefficients will reflect the influence of one 'unit' increase of that specific variable on the response outcome, the dependent variable. A positive beta coefficient will increase the response outcome, where a negative beta coefficient will decrease the response outcome. A larger beta coefficient, both negative or positive, will have a higher per unit impact on the outcome. The intercept and the corresponding beta coefficients of the applied model are given in Appendix H. The formula of the multiple linear regression function looks like Formula 4-4.

$$y = a + c_1 R_1 + c_2 R_2 + \cdots + c_n R_n + e \qquad \textbf{Formula 4-4}$$

In Formula 4-4 the e is the error term, also known as residual, which represents the margin of error within the statistical model, and n is the number of independent variables. In this equation is refers to the sum of the deviations within the regression line, so it provides an explanation for the difference between the results of the model and the actual observed results. The variables will be ranked according their beta coefficients in order to identify their relative importance. According to this ranking, the variables with the lowest coefficients will be dropped as a trade-off between data availability, model complexity and forecast accuracy.

Next to this, it is important to consider which performance measure to use, since typically there are many trade-offs between measures when optimizing a model. Improving the model on one performance measure will decrease the score on another (Amrit, Paauw, Aly, & Lavric, 2016). Which performance measure to choose depends amongst others on the desired level of interpretability or complexity or the bias/variance trade-off. That is why next to the R-squared both the MAPE and the Forecast Error are calculated, the results are shown in Table 16. As discussed in Section 3.10 the performance of the forecast is measured using the mean absolute percentage error and the forecast error. The results of these performance measures on the different models are given in Section 5.1 and 5.3, where the objective is to find the model that performance closely to the complex model but at least has an equal or better performance than the user forecasts.

Table 16: The result of the performance measures for the simplified models with different levels of LASSO's hyperparameter alpha.

Models:	# features used	R-squared prediction	MAE	MAPE	Forecast Error
Lasso(10)	13	NaN	1,017	15,529	13,033
Lasso(1)	19	0,42365	0,860	13,286	11,018
Lasso(0,1)	31	0,53700	0.811	12.424	10.393
Lasso(0,01)	41	0,58469	0,782	11,991	10,028
Lasso(0,001)	55	0,61541	0,762	11,690	9,755
Linear Regression	72	0,61775	0,760	11,676	9,744

Table 16 shows that an increase in alpha in the regularization method LASSO reduces the number of features used in the model. Also, with a decrease in the number of features used in the model the R-squared of the prediction is lower, which indicates that the level of variance explained by the model decreases. Besides, the other performances measures MAPE and forecast error both increase for each decrease in number of features used. Detailed analysis for each level of alpha are given in Appendix G.

The next section briefly explains the complex forecasting method to make the comparison between the simple model, the complex model and the user forecast in Chapter 5.

4.5.2 In comparison with the complex forecasting method

The forecasting method that is currently used by the company is defined by the users as a black box. Therefore, a brief explanation of this model is given. At the moment, a tree boosting method is used which is a widely used and effective machine learning method. The details of this complex method are explained in Chen & Guestrin (2016). However, (small) decision trees are simple models and are easy to interpret. A decision tree applies a stepwise method where the data is split in increasingly smaller branches. The output value of the branches is set to the mean of the true output of the samples in the branch. The explanatory variables are used as a decision rule for splitting the data. For example, whether a product has a second placement. As selection measure the branch with the highest increase to the performance measure is selected. This results in a greedy method, since the split is chosen where the current step is optimized instead of a split which might benefit the future tree. When using a decision tree the risk of overfitting can be limited by setting a minimum number of samples in each end node of the tree and restricting the number of levels (depth) of the tree. The minimum number of samples in an end node will prevent that each sample will get its own branch. Limiting the depth will cut off the final splits, which have a have no significant impact and therefore have a high chance to only contribute to overfitting. Because of this greedy approach, the decision tree model is highly dependent on the (number of) data it sees. These settings of minimum number of samples in an end node, minimum or maximum number of splits and leaves, are the hyperparameters of the model. The complex forecasting method is more sophisticated than the applied simple model because next to a different type of forecasting model is also optimizes the hyperparameters of the model by applying the grid search technique. Grid search builds a model for every combination of hyperparameters specified and evaluates each model, as you can image a wide set of options for each hyperparameter results in large computing time (Scikit-learn developers, 2018).

The complex method, which is called XGboost (Extreme Gradient boosting), is briefly explained in a simplified way using Chen & Guestrin (2016) and Pathak (2018). Boosting is a sequential techniques which works like an ensemble. As it combines multiple model outcomes based on the outcome of the previous instant. The outcomes predicted correctly are given a lower weight and those with mispredictions are weighted higher. The simplified idea behind boosting algorithms is to start with building a weak model, making conclusions about the various feature importance and hyperparameters, and then using those conclusions to build a new, stronger model and try to reduce the number of mispredictions. XGboost has integrated hyperparameters for amongst others, cross-validation and regularization which in the simple linear regression model have to be added seperately.

The complex forecasting method differs from the simple method because the complex method includes additional techniques, amongst others the extensive GridSearch, clustering of similar promotions, and a more sophisticated forecasting model XGboost, which is explained in Chen & Guestrin (2016) and Pathak (2018). Despite these differences, both methods use the steps of Section 4.5; data preparation, cross validation, fitting the model to the training data, thereafter predicting the outcome and evaluating the performance measures. Therefore, when using the same input data we can make comparisons between the simple and complex method on how well the methods are able to process this data into the output, measured by the performance measures.

4.6 Summary of the applied steps

The goal of this methodology is to determine which problem area of the root cause analysis to select and to identify how this area could be improved. Substantiated by a multi criteria decision making method by Saaty (1980), called the Analytical Hierarchy Process, from the areas: data availability, data quality and forecasting method, the data availability received the highest priority score. One purpose of this study is to assess the extent to which data availability, in the form of input parameters, is influencing the forecast performance measures and the acceptance of the forecast in business operations. Research question (F): “How to ensure that the tool is used and internalized?” is answered based on interviews with the users, concluding that the data availability could be improved by reducing the information overload and improving the user interface. Therefore, the aim of the methodology is to find the reduced number of variables for which the model still performs equal or better than the user forecasts. To reduce the number of variables the regularization method LASSO is applied with multiple levels for its hyperparameter alpha, which determines the level of variable reduction. The applied steps are summarized in an overview in Figure 26.

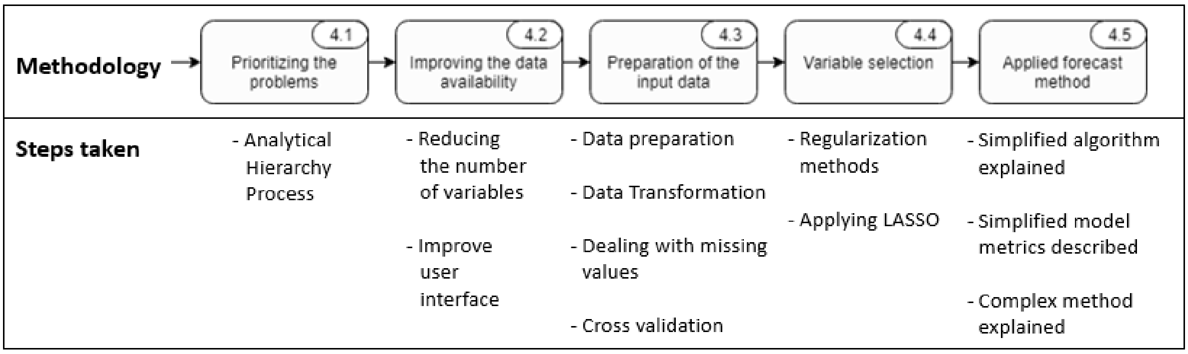


Figure 25: An overview of the applied steps in Methodology.

The output of the steps taken in each section in the methodology chapter (see Figure 26) is used as input for the analysis. The results of this analysis are presented in Chapter 5.

5. RESULTS

In this chapter the results of this study are presented. Starting in Section 5.1 to 5.3 with realization and validation of the simplified model. The simplified model is the result of the applied steps in Chapter 4. Subsequently, the deliverables of this study are presented: an improved user interface in Section 5.4 and a dashboard for monitoring data availability in Section 5.5. Finally, the simplified model is evaluated in Section 5.6, connecting the results in order to describe the impact of the realized forecasting method on the performance measures in business practice.

5.1 Realization of the simplified model

The aim of this section is come to a simplified forecasting model for product promotion. Requirement for a the simplified model is that it should at least have equal performance to the user forecasts.

Therefore, the results on the performance measures described in Section 4.5 are given in Table 17 and are used to compare the different model versions relative to the user forecasts. The alpha levels 10, 1, and 0.001 are removed from the results because of extreme values for at least one measure. The result of the performance measures of the simplified model (see two left columns in Table 17) are compared with the user forecast performance (see right columns in Table 17). The performance of both simple models is lower than the user forecast on 4 out the 6 measures. The goal is to maximize the level of R-squared while minimizing the levels of MAE, MAPE and forecast error. From Table 17 we see the user forecasts outperforms the simple models on R-squared, MAE, MAPE and Forecast error, but the simple models use less input variables and have lower missing values, 11% and 10% for simple(lasso0.1) and simple(lasso0.01) respectively, compared to 16% missing values for the user forecasts. Therefore, both simple models do not meet the criteria of equal performance compared to the user forecasts. However, the insights gained from these simple models e.g. the reduction in number of variables, is used as input for the complex forecasting methods.

Table 17: The results of the different forecasting models applied in the methodology (for explanation see Section 4.5).

Data results*	Simple (lasso 0.1)	Simple (lasso 0.01)	Complex (lasso0.1)	Complex (lasso0.01)	User
# variables	31	41	31	41	72
R-squared	0,74	0,74	0,77	0,88	0,87
MAE	4293,91	4294,98	3650,07	2841,86	3429,29
MAPE	659,12	660,57	128,37	125,22	136,60
Forecast Error	45,77	45,79	51,39	37,23	31,39
Missing values	11%	10%	11%	10%	16%

*Lasso alpha levels 10, 1, 0.001 and 0.0001 are removed wrong the results because of extreme values for at least one measure.

Using the output of the two versions of simple models, the reduced number of variables as input for the complex methods (explained in Section 4.5.2) results in respectively Complex(lasso0.1) and Complex(lasso0.01). The results of these versions of the complex models are presented in fourth and fifth column in Table 17. Furthermore, the performance measures of these two versions of the complex model are compared relative to the user forecasts in Table 18. Improvement on data availability is calculated by subtracting the percentage of missing values from Table 17 from 1, subsequently standardizing this value to the value of the current model. This results in the percentages improvement on data availability relative to the user (see Table 18). The color green indicates an improvement and the red color indicates an decrease of the performance relative to the objective of the table.

Table 18: Result of comparing the performance measures of two complex models to the user forecasts

Relative to User	Complex(lasso0.1)	Complex (lasso0.01)
# variables	57%	43%
R-squared	-12%	1%
MAE	-6%	17%
MAPE	6%	8%
Forecast Error	-64%	-19%
Improvement on data availability	6%	8%

Based on Table 18, Complex(lasso0.1) outperforms the user forecasts with 57% based on the number of variables included, also it outperforms the Complex(lasso0.01) on this measure (57% compared to 43% improvement compared to the user forecasts). However, on the five other measures the model Complex(lasso0.01) outperforms both the Complex(lasso0.1) and the user forecasts, see Table 18. Therefore, the model Complex(lasso0.01) is selected for subsequent analysis.

The impact of the differences in performance measures between the Complex(lasso0.01) and user forecasts on the demand volume forecasts is illustrated in Figure 26. With the probability density function, on the y-axis and the absolute difference between the prediction and the actual demand volume on the x-axis. It should be noted that the absolute difference results in a large tail with small density values, so for illustration purposes the quantity is limited to 10000. The circle 1 in Figure 27 indicate that the probability of a smaller absolute difference between the prediction and actual volume (<1000) at the model is higher than the user forecast. As a consequence, circle 2 in Figure 27 indicates that the user forecasts has more forecasts with a higher absolute difference between the prediction and actual volume.

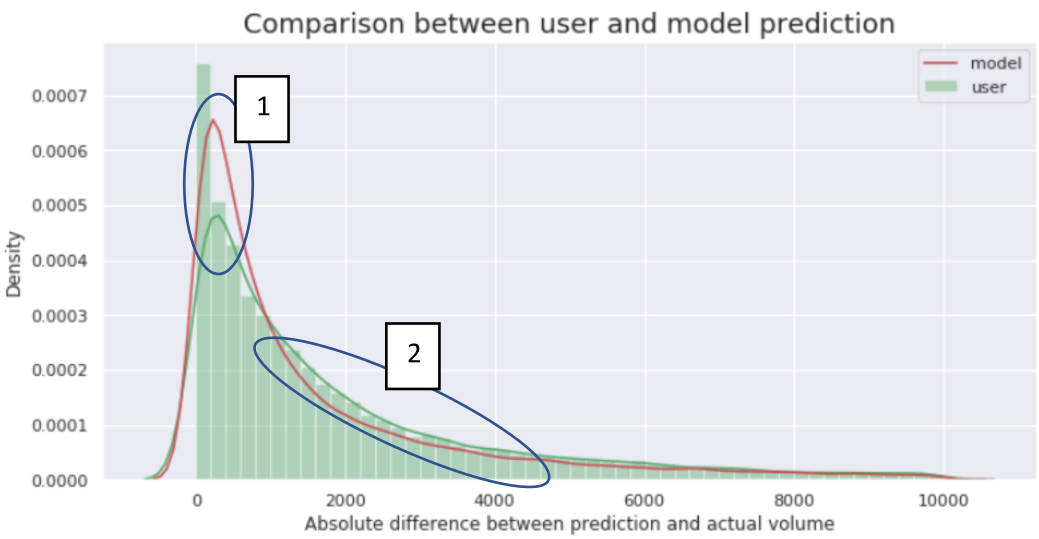


Figure 26: Complex(lasso0.01) compared with the user forecasts on the absolute difference between predicted and actual volume.

Based on the results presented in this section, the realized model of this study is, confusingly called, Complex(lasso0.01). To analyze the impact of this realized model on the current forecasting performance, in section 5.3 the Complex(lasso0.01) is compared with the current complex forecasting model. The variable reduction as a result of applying realized model is described in the next section.

5.2 Input variable reduction

The goal of this section is to determine which variables should be selected given the choice in the previous section applying Lasso with alpha 0.01, resulting in including 41 variables in the model. This determination is done using the relative feature importance, measured using the beta coefficients of the input variables. The list of the remaining variables is used as input for the next section, which explains how this list of most important input variables contributes to an increase in the forecast accuracy in business operation.

A snapshot of the output of the evaluation of the regression model is shown in Figure 28, to explain how this data provides insights in the contribution of the variables to the model. The complete output can be found in Appendix J. Figure 28 provides the least square estimates for each parameter, listed in the 'Coef' column next to the variable to which it corresponds. The calculated standard deviations are provided in the third column. The fourth columns 't' provides the test statistics. In linear regression, one wishes to test the significance of the parameters that are included. For any of the variables included in a multiple regression model, the null hypothesis (h_0) states that the coefficient β is equal to 0. The alternative hypothesis (h_1) may be one-sided or two sided, stating that β is either less than 0, greater than 0, or simply not equal to 0. The value of 't' for each input variable is calculated by dividing the values for 'coef' by the 'std err'. Then, the value follows from a t(n-p-1) distribution when p variables are included in the model and where n is the number of SKU's.

	coef	std err	t	P>|t|
const	-1.245e+04	3229.443	-3.854	0.000
base_price	-3009.4877	145.399	-20.698	0.000
baseline_units	3.2872	0.043	75.844	0.000
baseline_units_ext	0.1537	0.029	5.379	0.000
baseline_units_int	-0.1069	0.031	-3.478	0.001
baseline_vol	0.9240	0.036	25.863	0.000
discount_perc	150.4563	25.183	5.975	0.000
mechanism	-48.1423	10.653	-4.519	0.000
mechanism_type	521.5741	164.447	3.172	0.002
original_pid	0.5190	0.221	2.344	0.019
pid	-0.2591	0.156	-1.662	0.097
planned_discount_perc	123.3467	8.284	14.890	0.000
planned_promoted_price	72.2932	13.350	5.415	0.000
previous_promotion_week_distance	36.1266	50.552	0.715	0.475
prod_desc	-3.1153	0.414	-7.516	0.000
product_cu_per_sku	3.0745	19.086	0.161	0.872
product_dimension_13	1189.7601	137.224	8.670	0.000

Figure 27: Snapshot of the output results of the ordinary least square method to explain the insights that can be gained from these results.

The significance of the variables is explained using the variable highlighted in Figure 26 as an example. Consider this variable 'original_pid', the 'Coef' = 0.5190 and the 'Std err' = 0.221, resulting in t = (0.591/0.221) = 2.344. From the complete output result in Appendix J we find the degrees of freedom of this model is extremely large (>1000). We aim to include all variables that are significant with a confidence level op 95%. With the null hypothesis of the beta coefficients being 0 and the alternative hypothesis beta coefficient is not equal to 0, we use the two sided alpha. Looking at the t-distribution table (Moore, 1999) we find that with the degrees of freedom equal or greater than 1000 and a two sided alpha of 0.05, the critical value is 1.960. We reject the null hypothesis if the t-value is larger than this critical value. If the t-value is negative, we reject the null hypothesis if this t-value is smaller than the -critical value. In this example, the t-value of 'original_pid' is 2.344, which is larger than the critical value 1.960, therefore with a 95% confidence level we reject the null hypothesis. So, we say that the variable is significantly, with a 95% confidence level, different from 0, therefore contributes to forecasting the dependent variable 'total_units' and should be included in the model. Further analysis show us that an increase in the confidence level from 95% to 98% has no impact on the variable selection, the critical value with a two sided test with alpha = 0.02 is 2.326. Therefore, the 98%

confidence level is applied during the variable selection. Repeating these calculations for all variables in the model and selecting only those variables that significantly, with a 98% confidence level, contribute to the forecast of the 'total_units', results in 31 variables.

From this remaining list of 31 variables that contribute significantly to forecasting the dependent variable we aim to find their predictive strength. In other words, in order to find the most important variables we need to analyze their relative importance by standardizing the coefficients. Standardizing coefficients means that you can compare the relative importance of each coefficient in a regression model (Field, 2005). For example, let's say the model involved the quantity of the baseline and the promotional discount to forecast the total demand volume of the promotion. The baseline is measured in SKU's and the promotional discount as a percentage. Standardizing these variables means that they can be compared to each other in the model. The beta coefficients are calculated by subtracting the mean from the variable and dividing it by its standard deviation. The resulting standardized variables have a mean of zero and standard deviation of 1 (Field, 2005). The relative importance of the variables is calculated by putting the absolute value of the beta coefficients of each variable in descending order. Subsequently, divide the values of each beta coefficient with the largest beta coefficient to standardize the values to 1. As a result, the relative importance of the input variables are shown in Figure 29.

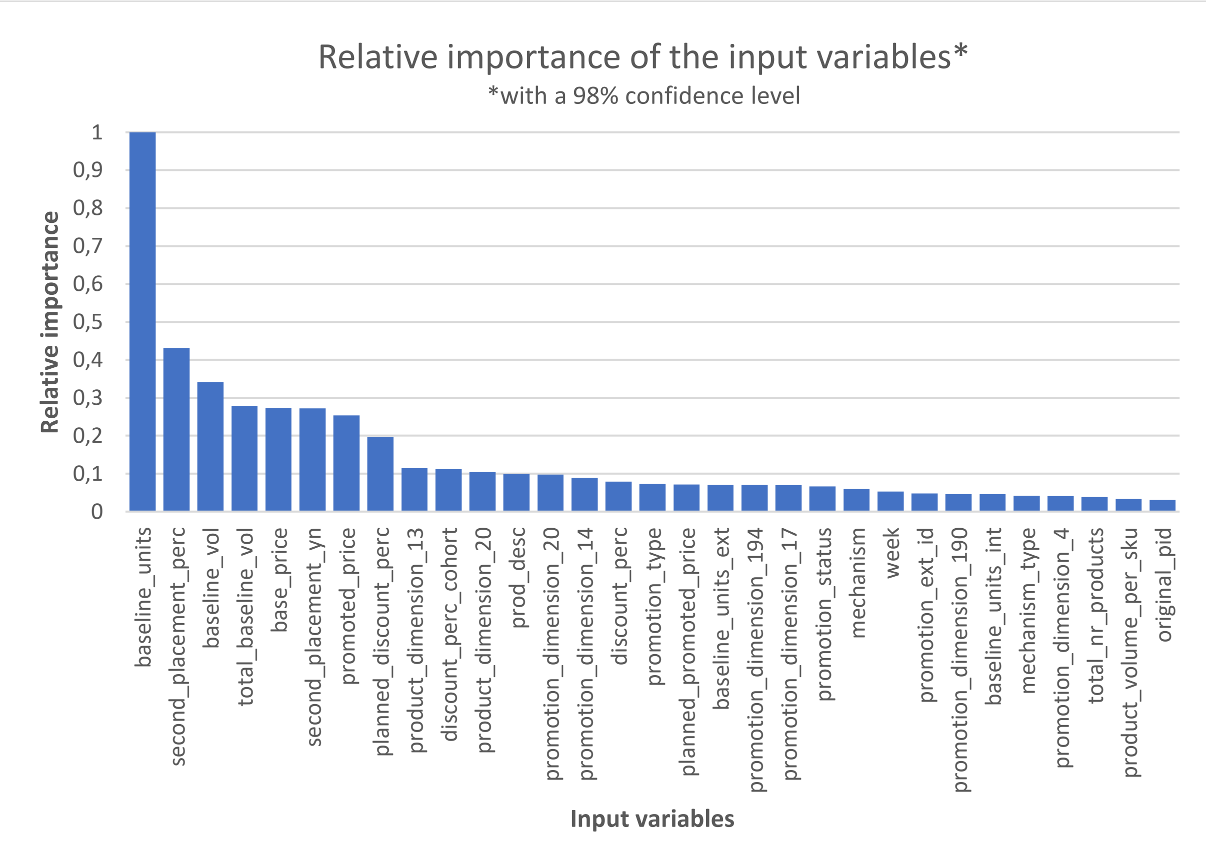


Figure 28: The relative importance of the input variables, with a 98% confidence interval, included in the model.

It must be stressed that this output of the model still requires human judgement. Human judgement is needed to check for correlation between the input variables. In Figure 29 we see for example, multiple input variables concerning the number of SKU's in the baseline. So, when translating the output of this variable selection (Figure 29) into the usable input variables for the user interface, human interaction is required to select only variables that have an unique contribution to the model.

The outcome of this variable selection after human judgement on correlation is presented in alphabetical order in Table 19. The list of input variables in Table 19 is used as input for the improved user interface described in Section 5.4 and for the introduced dashboard described in Section 5.5.

Table 19: The most important input variables after human judgement on correlation.

Variables			
Base price	Cluster	Mechanism	Promotion type
Baseline	Discount percentage	Productgroup	Second Placement
Brand	External ID	Promoted price	Total nr products
Category	Forecast Unit	Promotion status	Week

5.3 Realized model versus the current model

In section 5.1 the Complex(lasso0.01) model is selected as the best performing model relative to the user forecasts. In this section this selected model is compared with the current complex method on the same performance measures, see Table 20 for the results of this analysis.

Table 20: The result of comparing Complex(lasso0.01) and User forecasts relative to the Complex(current) model on the performance measures.

Relative to Complex (current)	Complex (lasso0.01)	User	Complex (current)
# variables	43%	0%	72
R-squared	-1%	-2%	0,89
MAE	-3%	-24%	2762,92
MAPE	0%	-9%	125,55
Forecast Error	-4%	13%	35,92
Improvement on data availability	8%	0%	16% missing values

In Table 20 the color green indicates an increase in performance of the model relative to Complex(current). The red color indicates a decrease. Table 19 shows a decrease in the number of variables by 43%, which corresponds with an improvement on data availability of 8% using the Complex(lasso0.01) instead of the Complex(current) model. However, this reduction in variables comes at a cost in the form of a decrease in the performance measures R-squared, MAE and forecast error. The MAPE is rounded to 0%, therefore assumed equal performances on this measure on aggregated level. The impact of this difference in performance measures on the forecast is shown in Appendix G, where the results of the forecasts for all three methods are shown.

In order to have a better understanding of the performance of the underlying promotions and SKU data next to the overall performance on aggregated level, analysis are performed on both cluster level and categorical level. For confidentiality reasons the outcome of the comparisons of between the current model, the realized model and the user forecasts is moved to Appendix L. Also, note that these values are adjusted for confidentiality reasons. However, looking at the average on categorical level between these three forecasts the results of the predictive models are within range of 5%, the user forecasts however have a higher deviation in the outcome.

5.4 Improved user interface

This section describes the improved user interface that is built using the insights on the most important input variables in Section 5.2.

Compared to the current user interface described in Section 2.5 the improved user interface contains the reduced number of variables, resulting in only the 16 important data input fields (Table 19) for the forecast of product promotions, see Figure 35**. By having all input fields in one overview screen the usability is increased. Relative to the existing situation of multiple separate workarounds in Excel, this improved user interface enables for consistency in timing and layout of the information by having all information in one place. Therefore, using the improved user interface might increase the internalization and use of the tool by the users. So it will facilitate the process from data to information, and from information to insights which is facilitated in the next section.

*** For confidentiality reasons the snapshot of the updated user interface is removed and located in Appendix M.*

5.5 Introduced dashboard for monitoring data availability

This section describes the dashboard for monitoring data availability that is introduced to the company as a deliverable of this study. As mentioned in Section 2.4, users lack knowledge of which data is used to establish the forecast of product promotions. The introduction of this dashboard makes it clear which variables are used and it allows for monitoring the data availability for each variable. This dashboard is already a proven success since it has already been viewed more than 600 times.

The introduction of this dashboard gives users the possibility to focus on specific variables instead of being overwhelmed with (big) data and limited information about numerous variables where the contribution to the forecast was indistinctive. To facilitate this focus and to be able to keep track of improvements on data availability a dashboard is build, see a Figure 30 for a snapshot. Also a worklist is provided (see Appendix E) to each specific user to show which variables within the planned promotions within their product portfolio are not filled in correctly. Besides checking if data is filled in for a variables this dashboard also allows for entering specific rule sets to ensure correct data is entered by setting boundaries or adding required data field.

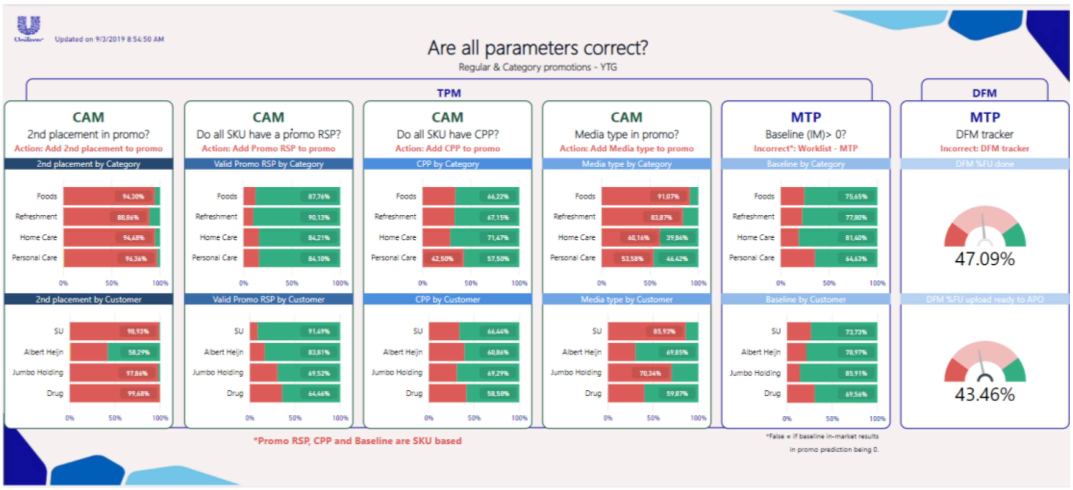


*Figure 29: Snapshot of the dashboard build to keep track of the data availability per variable**

**Please note that fictive data is used in this snapshot of the dashboard.*

As described in 3.2.3., the ability for diagnostics of the forecasting method is at least as important as the forecast accuracy. Simple model can be better tracked using e.g. dashboards. This dashboard improves the understanding by senior stakeholders. This leads to better coordination, which is accompanied by higher usability. Furthermore, the buy-in of users is increased due to the possibility to focus and coordination with senior stakeholders.

Let us assume that the data quality of the complex model and this model are equal. However, the data quality of future promotions is likely to be higher for those variables that will be tracked and focused on using the dashboard.

The dashboard is briefly explained, starting with the input of the important variables (see Figure 29) that users have to fill in are visualized in this dashboard. By visualizing the information users fill in it is possible to track and trace the progress. These input variables are translated into business practice, meaning that for example all variables relevant for explaining the 'consumer discount' in a promotion are combined to check if the 'consumer discount' is correctly filled in for the promotions. Consumer discount comprises of multiple variables (promo RSP, RSP, discount amount, discount percentage). This dashboard allows the viewer to adjust the aggregation level and therefore the level of detail one wants to view the information. The color green indicates the percentage of promotions that meet all criteria for that specific graph. Red indicates that at least one SKU within a promotion does not meet the requirements. For example, if at least one SKU within a promotion does not have a properly addressed consumer discount, the entire promotion will pop-up in red color. The promotions with incorrect data, when the requirement is not met, are presented on a work-list in a different tab within the dashboard. These work-list are made specifically for each type of user, each showing only relevant information for that user type. So, the MTP will get a worklist (Appendix E.2) with only baseline inputs they have to fix, while CAM see on their worklist (Appendix E.1) all sales related promotion variables.

5.6 Evaluation of the simplified model

In Section 5.1 to 5.3 the realization and validation of the simplified model is reported, in this section the simplified model is evaluated. This section describes whether or not an equal or higher performance on the operational forecast accuracy can be achieved using a simplified model with a higher data availability. The simple model using multiple linear regression performs worse than the user forecasts on the performance measures: MAPE, R-squared, MAE, Forecast Error. However, on the measures number of variables and data availability the simple models have a better score. Nevertheless, the insights on variable reduction obtained by building these simplified models are used as input for the complex model. Relative to the user forecasts the complex method using LASSO regularization with alpha 0.01 (Complex(lasso0.01)) has a higher performance on all six performance measures, except the forecast error. Therefore, the requirement that the applied forecast model must at least perform equal to the user forecasts is met.

Subsequently, the performance of the simplified model, Complex(lasso0.01) is compared to the current forecasting method. The result of this comparison on performance measures indicate that applying the Complex(lasso0.01) model would result in a decrease in performance on R-squared, MAE and Forecast Error of respectively 1%, 3%, and 4%, however an increase in performance based on the number of variables included and the improvement on data availability of respectively 43% and 8%.

As stated in Section 4.1, the focus is not on achieving the theoretical optimal forecast model, but focus on an entire forecast method which is applicable and accepted by the users. By lowering the number of features that are used in the model, and using a simplified model to involve users by revealing what they see as a 'black box', thereby increasing their buy-in for acceptance. Therefore, the results on data availability, measured in the improvement on missing values, and the forecast model, measured by the average of the difference in performance measures MAPE, R-squared, MAE, and forecast error between the current model and the realized model, are connected in order to give the result on the impact of the total forecasting method on the performance measure in practice, see Figure

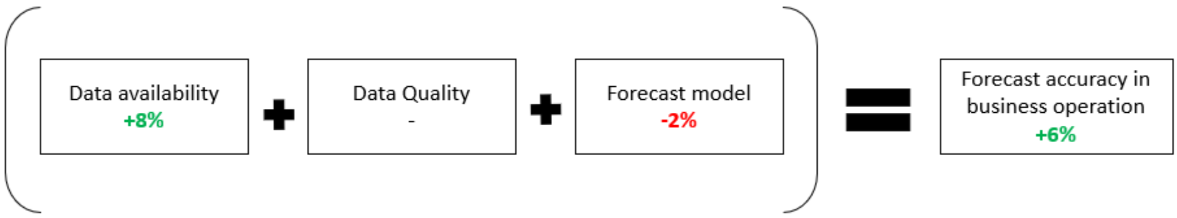


Figure 30: Result of the impact of the total forecasting method on the forecast accuracy in business operation.

Figure 31 shows that the implementation of the realized model leads to an increase in forecast accuracy in business operation of 6%, given that the level of data quality is kept constant.

6. CONCLUSION & RECOMMENDATIONS

This research is executed at the Integrated Operations department of Unilever Netherlands. The problem definition was: Volume forecast for product promotions are not accepted by users and therefore not used. Discussing the problems faced with forecasting promotional demand, has led to the following research question: "How to improve the current forecasting method for product promotions to ensure it is accepted and internalized by the users?".

Using the Analytical Hierarchy Process (Saaty, 1980) data availability is found to be significant given the priority score of 72% compared to 16% and 12% for respectively forecasting model and data quality. Therefore, the focus for improvement is on data availability rather than forecast model or data quality. In this study was found that important improvement steps are data availability, reducing information overload and improving the user interface.

As result of this study:

- **A simplified forecasting model is built with reduced number of input variables while still get a good performance of the forecast;**

Users do not accept the forecast because they see the current forecasting method as 'black box'. Therefore, a simplified model is built. Multiple linear regression is used as a starting point for the simplified model. This is substantiated using Van Loo (2006), Derks (2015), Divakar (2005), Leeflang, van Heerde & Wittink (2002), and Van den Heuvel (2009), who state that model relevance is primary to model sophistication. With this simplified model the users will be able to better understand the results of the significance and strength of the independent variables used to forecast the promotional demand more effectively. According to Fildes et al. 2009 the performance of the forecast should be measured using multiple measures. Substantiated by Hyndman (2014) and Field (2005) the measures MAPE, R-squared, MAE and the forecast error are used as performance measures in this study.

- **A modified user interface;**

The user interface is improved by only including the most important input variables in one overview instead of multiple forms and Excel sheets and the number of variables is reduced using the regularization methods LASSO and variable importance is calculated using the beta coefficients. With this modification to the user interface usability of the forecasting tool is improved because users will not be information overloaded.

- **A dashboard to monitor the data availability.**

Advantage of the simplified model is that monitoring and tracking progress of the forecast process is easier. Therefore a dashboard is built to monitor data availability. This dashboard improves the understanding by senior stakeholders. This leads to better coordination, which is accompanied by higher usability. Furthermore, the buy-in of users is increased due to the possibility to focus and coordination with senior stakeholders.

To conclude, it is proven (see Section 5.6) that a substantial improvement in forecast accuracy in business practice for Unilever can be obtained by improving the data availability of promotions.

So, in order to achieve this improvement it is recommended to the company to:

- **Implement the three improvements**; simplified model, improved user interface and a monitoring dashboard.

- **Perform periodic evaluation**
  With two perspectives:
  - Use and acceptance by users
  - Effectiveness of the model

Based on the outcome new adaptions can be done. This report with fundamental research can be used as reference for selecting appropriate measures. It is not a final phase because it doesn't ensure acceptance and internalization by the users yet, but it is a first step in continuous improvement process. It gives better preconditions because it 'overcomes' the major drawbacks of the current forecasting method.

7. DISCUSSION

The reliability of the quantitative research in this study is substantiated by using the same initial dataset containing the same SKU's and keeping the time period fixed of promotions that are included to cover the internal consistency. To ensure the test-retest reliability is met in the cross validation process the same random seed is used for all the forecasting models. Hence, the performance of the forecasting models can be compared because they use the same dataset split for training and test data.

In addition to the reliability, validity should be taken into account. Validity is the extent to which the scores from a measure represent the variables they are intended to (Price, Jhangiani, & Chiang, 2013). The scores for the pairwise comparisons as part of the Analytical Hierarchy Process are checked for consistency in Appendix B. The criterion validity, the extent to which people's scores on a measure are correlated with other variables (known as criteria) that one would expect them to be correlated with (Price et al. 2013). E.g. looking at the beta coefficients in Section 5.4 at for example the variable discount price, one would expect an increase in discount price would have a positive increase in the sales quantity. These criteria are checked and there is no reason to cast doubt on the validity of the measures.

The reader should bear in mind that the study is part of the continuous improvement. Therefore, the structure explained in Section 1.3 is used as a method that can be applied. The preliminary result in this cycle might change overtime, with the result that the largest improvement potential might shift from data availability to a different problem area. This shift in priority will imply that a different methodology needs to be applied, resulting in different analysis therefore different results.

The main weakness of this study was the paucity of measuring the acceptance of the users. Indicated as one of the most important factors, albeit not the most important factor, the level of acceptance is not made measurable. It is unfortunate that the study did not include a zero measurement of the level of acceptance at the beginning of this study and at the end of the study to analyze whether the level of acceptance is increased. Now, this study tried to translate potential improvement of the level of acceptance by removing the biggest frustrations of the users. However, it is hereby assumed that removing these frustrations implies that users will use the tool and acceptance the forecasts.

Despite the users frustration of the discrepancy between the forecasts on aggregate level and detailed level, the small sample size did not allow for in-depth analysis on detailed item level. The desired changes in effect on detailed item level require considerably more data. For example, the highest aggregation level, cluster level, might have significant data to forecast the demand for product promotions, however the detailed level, for a specific promotion within a certain product category might be based on 6 historical promotions. Logically following from this, the forecast on detailed level will not have the same accuracy as forecasts based on considerably more data points.

The generalizability of these results is subject to certain limitations. For instance, now only the data of promotions with promotion type 'Regular' and 'Category' are included. Also, the focus is determined to be on the largest retailer and with the largest promo pressure. When generalizing the findings of this research to different promotions types and different retailers other issues might arise in the forecasting method. It requires human judgement to identify which promotion types could be included in the current forecasting model. Recalibration of the parameters of the model might be required when generalizing these finding to different retailers or included different promotion types.

It is beyond the scope of this study to examine the quality of input, however the concept described as Garbage In is Garbage Out(GIGO) definitely applies to this study. Although, data cleaning is extensively done during this study it will not entirely solve the problem of data quality. Given the environment of a FMCG frequent changes in products and data will continue to occur. This requires a more structural approach. It is suggested to the company to further analysis this structural approach in the form of product life cycle management to ensure its data quality.

Looking at the results of the feature importance in Section 5.4 Figure 28, we see multiple variables related to the baseline with a high relative importance. However, the focus of this research is on the input variables for promotional demand, rather than regular (base) demand. As this variables turns out to be one of the most important variables it is recommended to the company to broaden the scope by including the baseline forecast. The performance on forecast accuracy is highly dependent on the quality of the baseline forecast, because the total demand quantity is determined by the baseline times the lift factor.

7.1 Limitations

Since the forecasting models are based on historical data, the inherent effect is that historical demand is expected to be a good representation of future behavior. If substantial changes occur in e.g. the shoppers perception on products, this should be evaluated and recalibration of the parameters might be desired.

The current research focuses solely on improving the forecast of product promotions. From the company perspective, decisions should be made based on the total sales of both base and promotional demand. In order to reach company objectives, neither only regular sales nor promotional sales should be optimized.

The simplified model is built under the assumption of the normal distribution of the data, this has the following properties: Observations around the average are the most likely to occur, the more values deviate from the mean, the less likely it is to observe these values, and values above and below average are equally likely. Caution should be applied when these assumptions are used into business practice. When looking at the distribution histogram of the total units against the frequency, it is questionable that promotions with a small volume are equally important for the company as large volume promotions with an equal distance to the mean. Also, although extremely large promotions do not occur frequently, they are important for the company in a business strategic way.

Furthermore, in order to get significant estimates of the beta coefficients of the regression model the data must meet certain conditions (Field, 2005). Therefore, the following assumptions should be validated when fitting a (multiple) linear regression model to the data:
1. Non linearity of the response-predictor relationships
2. Correlation of error terms
3. Non-constant variance of error terms
4. Outliers
5. High leverage points
6. Collinearity

According to James et al. (2013), identifying and overcoming these problems is as much an art as a science. One of the most important assumptions when determining the importance of variables in this study is multicollinearity.

Multicollinearity
Multicollinearity exits when there is a strong correlation between two or more independent variables in the regression model (Field, 2005). This poses a problem if there is perfect collinearity, when two independent variables have a correlation coefficient of 1. If there is perfect collinearity between independent variables it becomes impossible to obtain unique estimates of the regression beta coefficients because there are an infinite number of combinations of coefficients that would work equally well (Field, 2005).

Having uncorrelated independent variables is beneficial because in that way the increase in independent variables will contribute to the level of variance explained by the model, resulting in an increase of R-squared. Next to this, multicollinearity makes it difficult to assess the individual importance of independent variables. To put it in other words, if two independent variables are highly correlated, and each accounts for the similar variance in the outcome, then how can we determine which of these two variables is important? Simply, we can’t tell, both variables are interchangeable.

Multicollinearity can be checked using a correlation matrix to see which variables are highly correlated (>0.90). Next to this, Minitab provides collinearity diagnostics, one of which is the variance inflation factor (VIF). The VIF indicates whether an independent variables has a strong linear relationship with other independent variables. If the average VIF > 1, multicollinearity may be biasing the regression model (Field, 2005).

7.2 Recommendations for further research

This sections describes recommendations for further research. It is suggested to investigate the potential of using other forecasting methods, to overcome the problem of multicollinearity. The rising popularity, use and development of data mining techniques such as neural networks might be promising areas to look into. More sophisticated methods have not yet necessarily shown to be more accurate (Makridakis et al., 2018). However, strong development in this area might provide new insights. An important aspect to take into account for more sophisticated models is the understandability for practitioners must be ensured. To be used in practice, it is important that these sophisticated models are understood. Also, it is advised to investigate which input variables could be automatically filled to increase the data quality and reduce the dependency on user input.

Another suggestion for further research, based on Brownlee (2018), is to investigate the opportunity to reduce the variance in the forecasting model by exploring the options to ensemble the predictions, ensemble the parameters or to find a way to increase the size of the data set to train the model.

Besides, confidence interval levels of the forecast would be a fruitful area for further work. When the sample size, mean, and variance of a forecast are available it is possible to incorporate a confidence interval in the forecast. This confidence interval might help the users to indicate a range in demand quantity for their forecasts. In the long-term, this could be used to pro-actively plan future promotion campaigns.

8. REFERENCES

Adefulu, A.D. (2015). Promotional Strategy Impacts on Organizational Market Share and Profitability. Acta Universitatis Danubius. Economica,11 (6).

Ali,O., Sayin, S., Van Woensel, T. &Fransoo, J. (2009). SKU demand forecasting in the presence of promotions. Expert System with applications, 36: 12340-12348.

Alvarado-Valencia, J. Barrero, L.H., Onkal, D. &Dennerlein, J.T. (2016). Expertise,credibility of system forecasts and integration methods in judgmental demand forecasting. Journal of Forecasting. http://dx.doi.org/10.1016/j.ijforecast.2015.12.010.

Amrit, C., Paauw, T., Aly, R. & Lavric, M. (2016). Identifying child abuse through text mining and machine learning. Expert Systems With Applications, 88: 402–418.

Armstrong, J.S. (2001). Principles of forecasting: A Handbook for Researchers and Practitioners. Philadelphia: University of Pennsylvania.

Armstrong, J. S. (2012).Illusions in Regression Analysis. International Journal of Forecasting, 28: 689 - 694.

Box, G. E. P. (1976), "Science and statistics" (PDF), Journal of the American Statistical Association, 71: 791–799, doi:10.1080/01621459.1976.10480949.

Blattberg, R.C. & Hoch, S.J. (1990). Database Models and Managerial Intuition: 50% Model + 50% Manager. *Management Science 36*(8):887-899. http://dx.doi.org/10.1287/mnsc.36.8.887

Brockwell, P.J. & Davis, R.A. (2016). *Introduction to Time Series and Forecasting*. Berlin: Springer. Doi: 101007.

Brownlee, J. (2018). How to reduce variance in a final machine learning model. Machine Learning Process. https://machinelearningmastery.com/how-to-reduce-model-variance/.

Chaharsoughi, S.A. & Yasory, T. H. (2012). Effect of sales promotion as a tool on customer attention to purchase: A case study of auto maker company. *African Journal of Business Management, 6*(5): 2007-2014.

Chambers, J.C., Mullick, S.K. & Smith, D.D. (1971). How to Choose the Right Forecasting Technique. *Harvard Business Review, 7.*

Chen, T., Singh, S., Taskar, B. & Gyestin, C. (2015). Efficient second-order gradient boosting for conditional random fields. In Proceeding of 18th Artificial Intelligence and Statistics Conference (AISTATS'15), 1.

Chopra, S. & Meindl, P. (2001). Supply Chain Management: Strategy, Planning, and Operation. New Jersey: Prentice Hall.

Cooper, L.G., Baron. P., Levy, W., Swisher, M. & Gogs, P. (1999). Promocast: A new forecasting method for promotion planning. *Marketing Science 18*(3): 301-316. http://dx.doi.org/10.1287/mksc.18.3.301.

Corsten, D. & Gruen, T. (2003). Desperately seeking shelf availability: an examination of the extent, the causes, and the efforts to address retail out-of-stocks. *International Journal of Retail & Distribution Management, ISNN:* 0959-0552.

Currie, C. S. and Rowley, I. T. (2010). Consumer behaviour and sales forecast accuracy: What's going on and how should revenue managers respond & quest. Journal of Revenue & Pricing Management, 9(4): 374-376.

Derks, L.A.D. (2015). *Improving promotion forecasts for the Dutch FMCG market a research executed at Royal FrieslandCampina N.V.* (master's thesis). Eindhoven University of Technology.

Dietvorst, J.D., Simmons, J.P. & Massey, C. (2014). Algorithm Aversion: People Erooneously Avoid Alforithms after seeing them err. Journal of Experimental Psychology. *University of Pennsylvania.* http://dx.doi.org/10.1037/xge0000033.

Divakar, S., Brian T. Ratchford, B.T. & Venkatesh, S. (2005) Practice Prize Article—CHAN4CAST: A Multichannel, Multiregion Sales Forecasting Model and Decision Support System for Consumer Packaged Goods. *Marketing Science 24*(3):334-350. http:// dx.doi.org/10.1287/mksc.1050.0135.

Fabricant, R. (2013) Article on User Experience from Harvard Business Review. Retrieved at 10-07-2019 from: https://hbr.org/2013/01/scaling-your-ux-strategy

Fildes, R., Goodwin, P., Lawrence, M. & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments. An empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting, 25* (1): 3-23.

Field, A. (2005). *Discovering Statistics Using SPSS.* Dubai: Oriental Press.

García, S., Luengo, J. & Herrera, F. (2015). *Data Preprocessing in Data Mining.* New York: Springer.

Green, S. (1991). How many subjects does it take to do a regression analysis. *Multivariate behavioral reserach*, 499-510.

Heerde Van, H.J., Gupta, S. & Wittink, D.R. (2002). Sales Promotion Bump Due to Brand Switching. *Journal of Marketing Research. 10:* 481–49

Heerkens, H., & Van Winden, A. (2017). *Solving Managerial Problems Systematically*. Retrieved 4 16, 2019, from https://research.utwente.nl/en/publications/solving-managerial-problems-systematically

Heuvel Van, F.P. (2009). Action products at Jan Linders Supermarkets.

Herrala, O. (2018). Evaluating cannibalizaton between items in retail promotions. (master's thesis). Aalto University

Hyndman, R.J. & Kostenko, A.V. (2007). Minimum sample size requirements for seaonal forecasting models. *Foresigt, 6.*

Hyndman, R.J. (2014). Measuring forecast accuracy. *Journal of Forecasting 22*(4), 279-688.

James, G., Hastie, T., Witten, D. & Tibshirani, R. (2013). *An Introduction to Statistical Learning. With application in R.* New York: Springer.

Kerkkänen, A., Korpela, J. & Huiskonen, J. (2008). Demand forecasting errors in industrial context: Measurement and impacts. Production Economics 118: 43-48. doi:10.1016/j.ijpe.2008.08.008

Kim, W., Byoung-Ju, C., Eui-Kyung, K. & Doheon, L. (2001). A taxonomy of dirty data. *Data Mining and Knowledge Discovery, 7:* 81–99.

Kock, J. (2002). Improving the promotion forecast accuracy at Coca-Cola enterprises. (Masterthesis's). Eindhoven University of Technology.

Lawrence, M., Goodwin, P., O'Connor, M. & Önkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. International Journal of Forecasting, 22 (3): 493-518. https://doi.org/10.1016/j.ijforecast.2006.03.007.

Loo Van, M. (2006). *Out-of-Stock reductie van actieartikelen, Model voor vraagvoorspelling en logistieke aansturing van actieartikelen bij Schuitema/C1000*. Master Thesis, Eindhoven University of Technology, Eindhoven.

Lee, Hau & Padmanabhan, V. & Whang, Seungjin. (1997). Information Distortion in a Supply Chain: The Bullwhip Effect. Management Science. 50. 1875-1886. 10.1287/mnsc.43.4.546.

Macaira Medina, P., Thomé Tavares, M. A., Oliveira, F.L.C. & Ferrer, A.L.C. (2018). Time series analysis with explanatory variables: *A systematic literature review. Enviromental Modelling & Software, 107:* 199-209.

Makridakis, S., Wheelwright, S.C. & Hyndman, R.J. (1998*). Forecasting: Methods and Applications*. Hoboken: John Wiley and Sons. Reviewed by: Alvaro E. Faria, Jr. (2002). Lancaster University; published in the International Journal of Forecasting, 18: 158-159.

Makridakis, S., Spiliotis, E. & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward*. PLoS ONE 13*(3). :https://doi.org/10.1371/journal.pone.0194889

McCarthy, I.P., Tsinopoulos, C., Allen, P. & Rose-Anderssen, C. (2006). New Product Development as a Complex Adaptive System of Decisions. *The Journal of Product Innovation Management, 23:*437–456.

Mentzer, J.T. & Moon, M.A. (2005). *Sales forecasting management: A demand management approach*. London: Sage Publications, Thousand Oaks.

Moore, R. (1999). Critical values for the t-distribution. Mathematics Department, Macquarie University, Sydney. https://faculty.washington.edu/heagerty/Books/Biostatistics/TABLES/t-Tables/

Müller, A.C. & Guido, S. (2017*). Introduction to Machine Learning with Python: A Guide for Data Scientists.* Sebastopol: O'Reilly Media.

Nielsen (w.d.). Business facts & figures. External marketing database www.nielsen.com

O'Connor, M. & Webby, R. (1996). Judgemental and statistical time series forecasting: a review of the literature*. International Journal of Forecasting, 12* (1): 91-118. https://doi.org/10.1016/0169-2070(95)00644-3.

Pathak, M. (2018). Using XGboost in Python. DataCamp. Retrieved from https://www.datacamp.com/community/tutorials/xgboost-in-python

Peters, J. (2012). Improving the promotional forecasting accuracy for perishable items at Sligro Food Group B.V. (Masterthesis's). Eindhoven University of Technology.

Price, P.C., Jhangiani, R., & Chiang, I.A. (2013). Research methods in psychology: Reliability and Validity of Measurement. *The Slayer Foundation*

Unilever. (w.d.). Business facts & figures. Retrieved from https://www.unilever.com/about/who-we-are/about-Unilever/

Raghubir, P., Inman, J. & Grande, H. (2004). The Three Faces of Consumer Promotions. *California Management Review, 46*(4): 23-42. https://doi.org/10.2307/41166273.

Ramanathan, U. & Myldermans, L. (2010). Identifying the underlying structure of demand during promotions: A structural equation modelling appraoch. *Expert system with applications, 38:* 5544- 5522.

Raschka, S. (2015). *Python Machine Learning. Unlock Deeper Insights into Machine Learning With This Vital Guide to Cutting-edge Predictive Analytics.* Birhmingham: Packt Publishing Limited.

Ries, E. (2011). *The Lean Start Up. How Relentless Change Creates Radically Successful Businesses.* London: Penguin Books.

Royston, P. (2004). Multiple imputation of missing values. *The Stata Journal, 4*(3): 227-241.

Saaty, T.L. (1980). *The Analytic Hierarchy Process*. New York: McGraw-Hill.

Schrijver, B.G.F. (2009). Forecasting for promotion items at Metro Cash & Carry Netherlands. (Masterthesis's). Eindhoven University of Technology

Scikit-learn developers (2018). *1. Supervised learning*. Retrieved from https://scikit-learn.org/stable/supervised_learning.html#supervised-learning.

Shankar, V., Inman, J. J., Mantrala, M., Kelley, E., & Rizley, R. (2011). *Innovations in Shopper Marketing: Current Insights and Future Research Issues. Journal of Retailing, 87, S29–S42.* doi:10.1016/j.jretai.2011.04.007

Syntetos, A., Babai, M., Dallery, Y. & Teunter, R. H. (2009). Periodic control of intermittent demand items: theory and empirical analysis. *Journal of the Operational Research Society. 60:* 611 – 618.

Syntetos, A. A., Babai, Z. Boylan, J.E., Kolassa, S. & Nikolopoulos, K. (2015, 2016) Supply chain forecasting: Theory, practice, their gap and the future. European Journal of Operational Research, 1-26.

Teng, C.M. (1999). *Correcting Noisy Data.* School of Computer Science and Engineering University of New South Wale.

Trapero, J.R., Pedregal, D.J., Fildes, R. & Kourentzes, N. (2013). Analysis of judgmental adjustments in the presence of promotions. *International Journal of Forecasting, 29*(2): 234-243. https://doi.org/10.1016/j.ijforecast.2012.10.002.

Trusov, M., Bodapati, A.V. & Cooper, L.G. (2006). Retailer promotion planning: improving forecast accuracy and interpretability. Journal of Interactive marketing 20: 3-4.

Tunikova, O. (2018). Article on information overload from Medium. Retrieved at 17-07-2019 from: https://medium.com/@tunikova_k/are-we-consuming-too-much-information-b68f62500089

Van Donselaar, K., Van Woensel, T., Broekmeulen, R. & Fransoo. (2006). Inventory control of perishables in supermarkets. *Production Economics 104:* 462-472.

Van der Poel, T. (2010). *Improving the promotion forecasting accuracy at Unilever Netherlands.* (master's thesis). Eindhoven University of Technology.

Wittink D., Leeflang, P. & Van Heerde, H. (2002). How Promotions Work: SCAN*PRO-Based. Evolutionary Model Building. *Schmalenbach Business Review, 54:* 198–220.

Yelland, P.M. (2010). Bayesian forecasting of parts demand*. International Journal of Forecasting, 26*: 374-396.

Zionts, S. (1979). *MCDM—If Not a Roman Numeral, Then What?.* Interfaces 9(4):94-101. http://dx.doi.org/10.1287/ inte.9.4.94.

APPENDICES

Appendix A – Unilever Brands

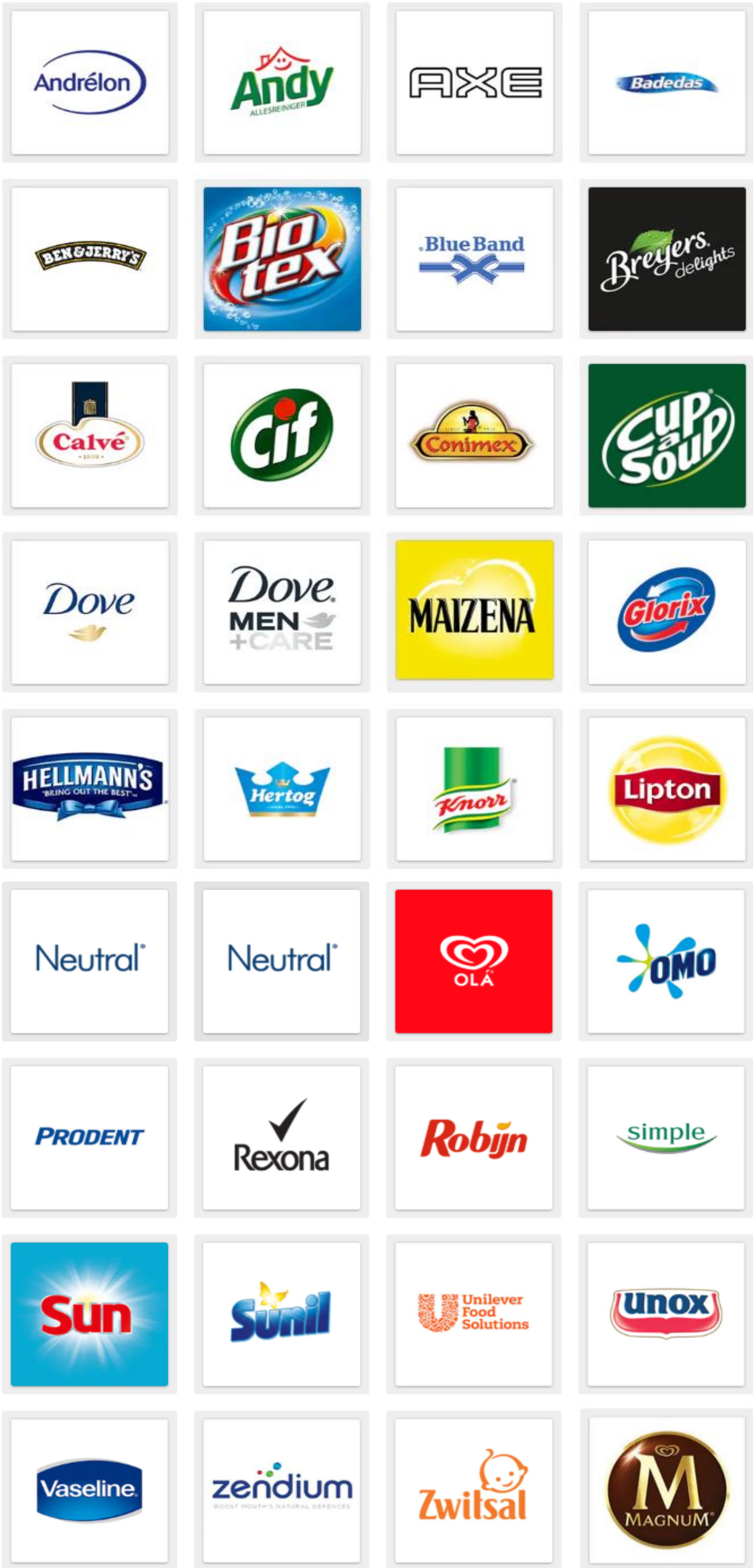


Figure 31: Overview of the Unilever brands available in the Netherlands (www.unilever.com)

Appendix B – Analytical Hierarchy Process

In this section the calculations of the Analytical Hierarchy Process are given. The following steps are repeated for Table A to Table D. The normalized values of the pairwise comparisons are calculated by for each factor (i and j) by a_{ij} / sum (over j) for a_i to normalize the weighting criteria. The priority vector is calculated by the sum of that factor divided by the number of criteria n. E.g. 2,30 / 4 = 57,39% in Table A for criterion A. The largest eigen value is computed in order to check for the consistency of the decision maker, see Formula B-1.

$$\lambda_{max} = \sum_{i=1}^{n}\left[\left(\sum_{j=1}^{n} a_{ij}\right) * w_i\right] \qquad \textbf{Formula B-1}$$

For this check for consistency both the consistency index (CI), see Formula B-2 and consistency ratio (CR), see Formula B-3 are calculated. The decision is found consistent if the consistency ratio (Formula B-3) is less than 10%. The Random Consistency Index (RCI) is retrieved from Table B.2, which is adapted from Saaty (1980).

$$Consistency\ Index\ (CI) = \frac{\lambda_{max} - n}{n-1} \qquad \textbf{Formula B-2}$$

$$Consistency\ Ratio = \frac{Consistency\ Index}{Random\ Consistency\ Index} \qquad \textbf{Formula B-3}$$

Table 2: The Random Consistency Index (RCI)

n	3	4	5	6	7	8	9	10	11	12	13	14	15
RCI	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	1.51	1.48	1.56	1.57	1.59

Table A: The normalized values of the analytical hierarchy process to calculate the priority vector for the criteria.

Normalized weighting criteria		A	B	C	D	Sum	Priority Vector
Influenceability	**A**	0,63	0,69	0,53	0,45	2,30	57,39%
Impact on FC accuracy	**B**	0,21	0,23	0,38	0,35	1,17	29,13%
Ease of understanding	**C**	0,09	0,05	0,08	0,15	0,36	9,03%
Complexity	**D**	0,07	0,03	0,03	0,05	0,18	4,45%
	Sum	1,00	1,00	1,00	1,00	4,00	1,00

lambda(max)	4,2692	CI	0,08972	
		CR	9,97%	<10% then consistent
		RCI	0,9	if matrix n is 4

Tabel B: The normalized values of the analytical hierarchy process to calculate the priority vector for the factor influenceability.

Normalized with respect to the factor Influenceability		X	Y	Z	Sum	Priority Vector
Data availability	**X**	0,74	0,64	0,79	2,17	72,35%
Data quality	**Y**	0,11	0,09	0,05	0,25	8,33%
Forecasting method	**Z**	0,15	0,27	0,16	0,58	19,32%
	Sum	1,00	1,00	1,00	3,00	100,00%

lambda(max)	3,1115	CI	0,055731851	
		CR	9,61%	<10% then consistent
		RCI	0,58	if matrix n is 3

Tabel C: The normalized values of the analytical hierarchy process to calculate the priority vector for the factor forecast accuracy.

Normalized with respect to Impact on FC accuracy		X	Y	Z	Sum	Priority Vector
Data availability	**X**	0,74	0,79	0,64	2,17	72,35%
Data quality	**Y**	0,15	0,16	0,27	0,58	19,32%
Forecasting method	**Z**	0,11	0,05	0,09	0,25	8,33%
	Sum	1,00	1,00	1,00	3,00	100,00%

lambda(max)	3,1115	CI	0,055731851	
		CR	9,61%	<10% then consistent
		RCI	0,58	if matrix n is 3

Tabel D: The result of the prioritization using the Analytical Hierarcy Process

	A	B	Priority factor
Adjusted weight factor	*0,663*	*0,337*	
Data availability	0,480	0,244	72,35%
Data quality	0,055	0,065	12,03%
Forecasting method	0,128	0,028	15,62%

CI	0,145449478	9,83%	<10% then acceptable
CR	1,48		
RCI	0,58		

Appendix C – Interview with users

Interviews with the users

The semi-structured interview is conducted in the beginning phase of this study with 13 users who are representatives from the different customer sales teams and planning teams. The results of the interviews are combined into three bins (see Figure 32): issues related to the forecasting tool, gap of knowledge, or business related questions. The interpretation of these results is given in Section 4.2.

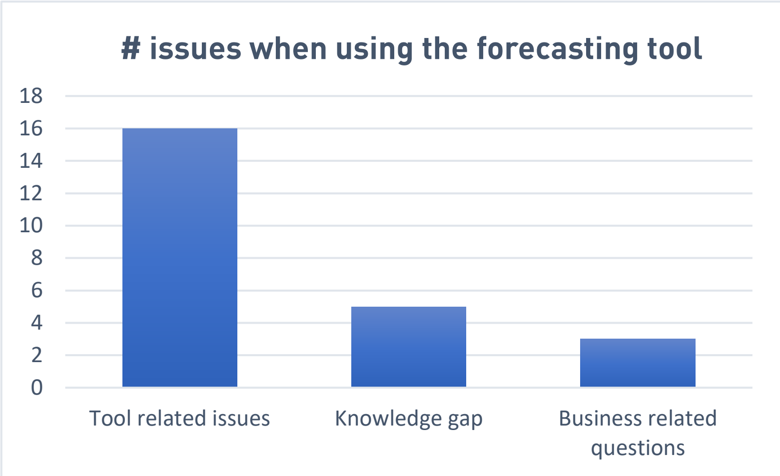
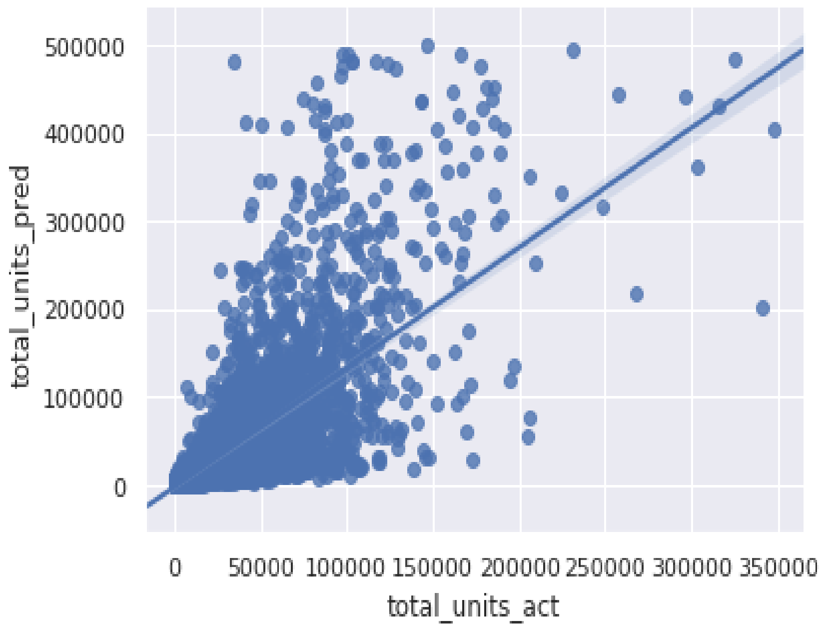


Figure 32: Result of the interviews with multiple users regarding issues when using the current forecasting tool.

Appendix D – Comparison between simple and complex model

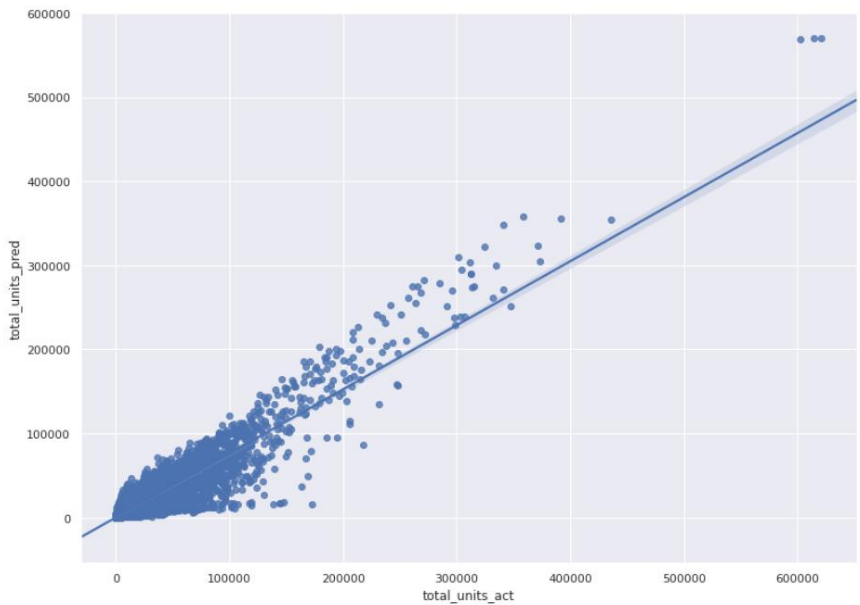
In this section comparisons are made between the forecasts of the total units by the simplified model (Figure A), the complex model with the reduced number of variables as input (Figure B), the complex model (Figure C) and by the users (Figure D). Based on these comparison the complex model outperformances both the simplified model and the users on forecasting the total units of a product promotions.

Forecast using the simple linear regression model with reduced variables.



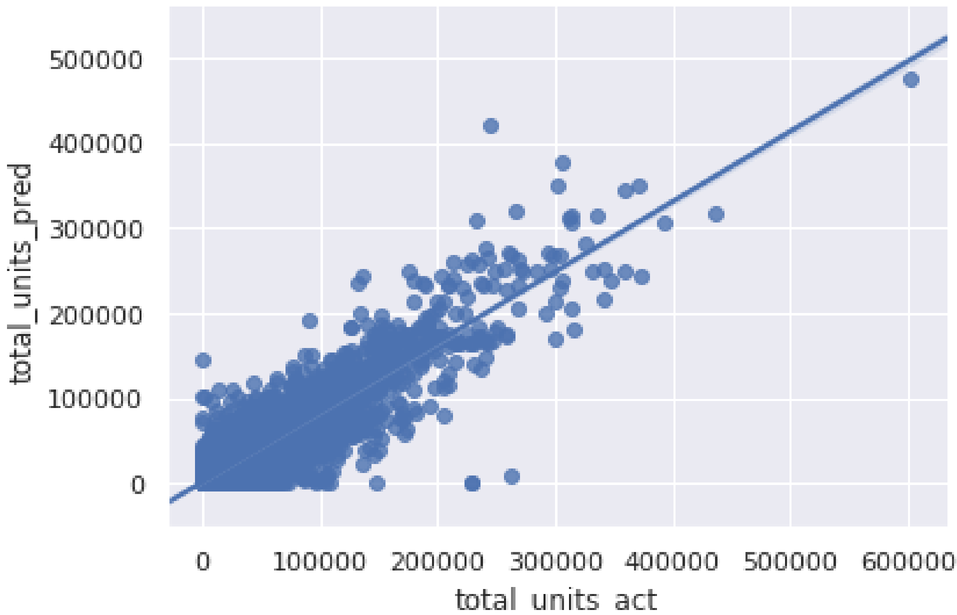
Figuur A: Forecast of the total units using the simple linear regression.

Forecast using the complex model using the reduced number of variables by LASSO with alpha 0.01, called Complex(lasso0.01).



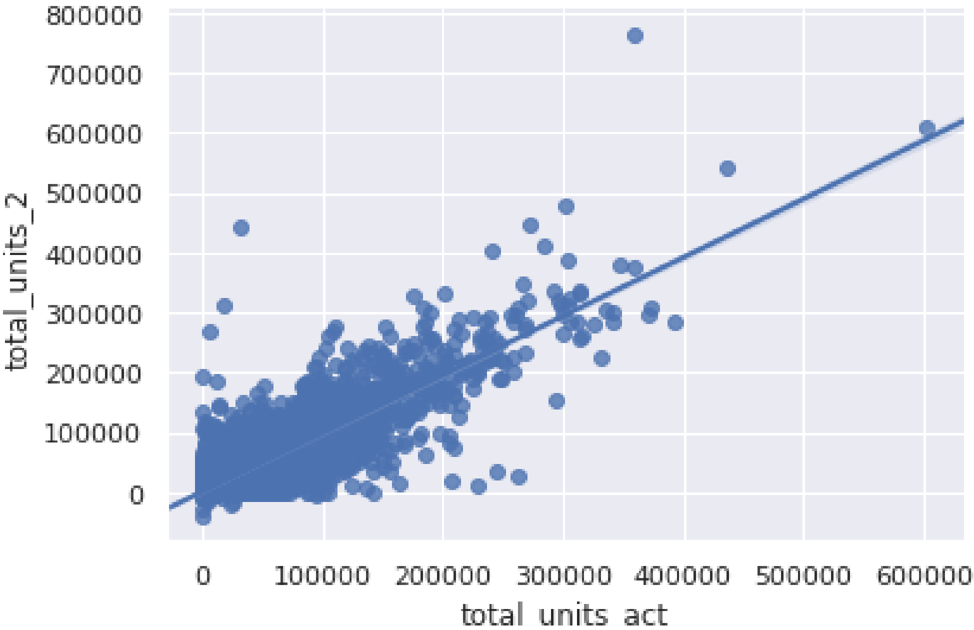
Figuur B: Forecast of the total units using the Complex(lasso0.01).

Forecast using the current complex method, called Complex(current).



Figuur C: The forecast of the total units using the complex model

Forecast by the users, called user forecast.



Figuur D: Forecast of the total units by the users

Appendix E – Dashboard snapshots

In this section multiple snapshots of the parameter tracker dashboard built for this research are presented. Two worklists for both the CAM and the MTP are shown. These worklists will help the users to focus when solving issues, enables users and management to track progress when making adjustments.

Appendix E.1 Worklist – CAM

This worklist helps the CAM to pin point which promotions need their attention instead of having to go through the whole assortment and promotions they manage. Using this worklist the CAM can directly spot which rule is violated, show as red dot and 'FALSE'. For example, if there is no SKU on 2nd placement in a promotion, that promotion will show up on this worklist (see Figure 33)

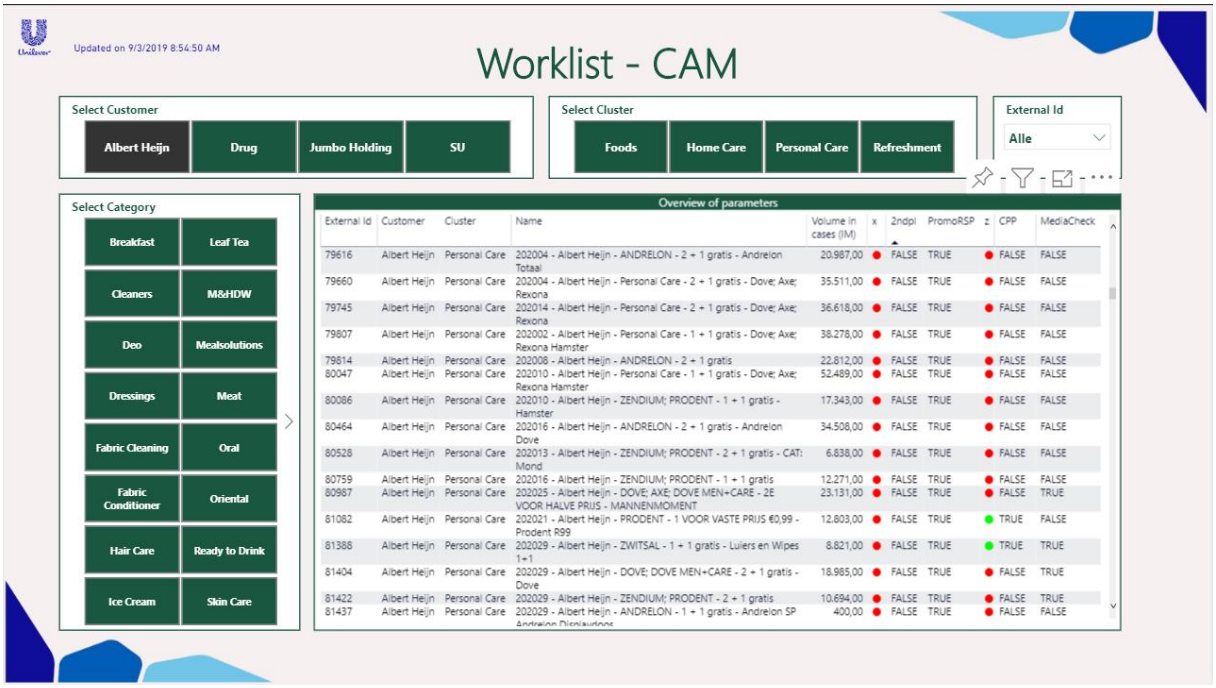


Figure 33: Snapshot of 'Worklist - CAM' from the parameter tracker dashboard.

Appendix E.2 Worklist - MTP

Worklist for the Mid Term Planners (see Figure 34) to check the SKU's with zero or negative baseline. Zero baseline might indicate that a SKU is planned by the CAM but no longer in production, therefore it should be taken out of the assortment. A negative baseline volume indicate an erroneous value and should be corrected.

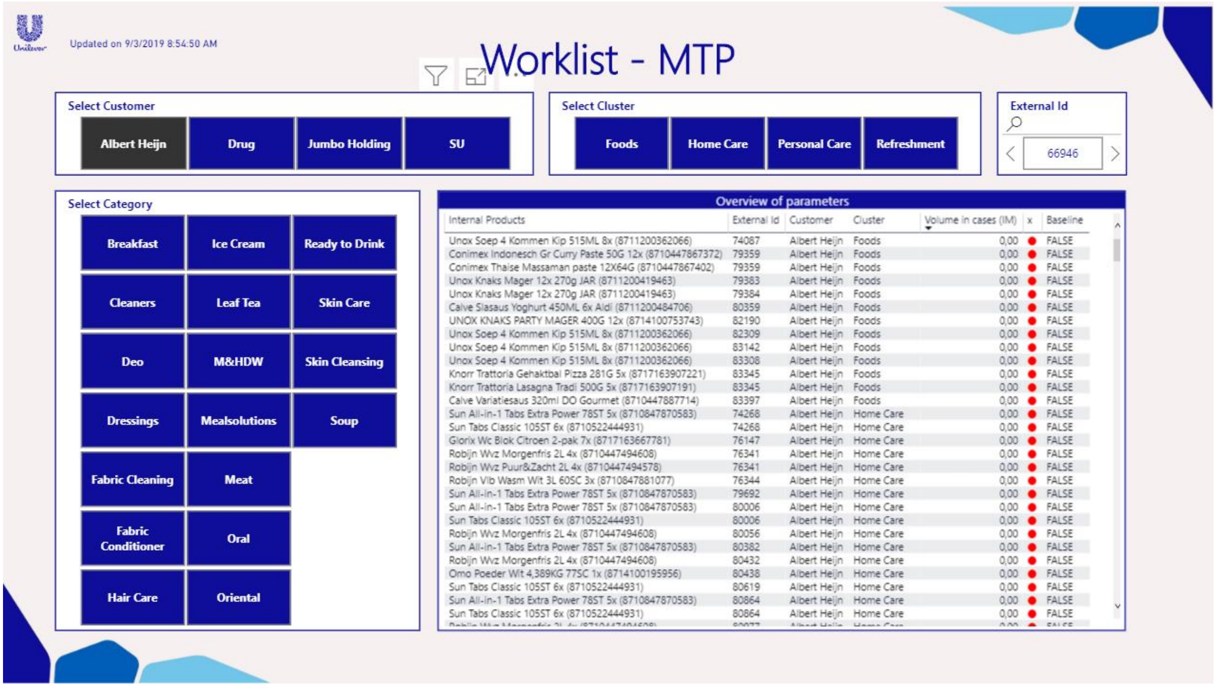


Figure 34: Snapshot of 'Worklist - MTP' from the parameter tracker dashboard

Appendix F – Flowchart of the realized forecasting method

This section present the flowchart of the realized forecasting method (see Figure 35).

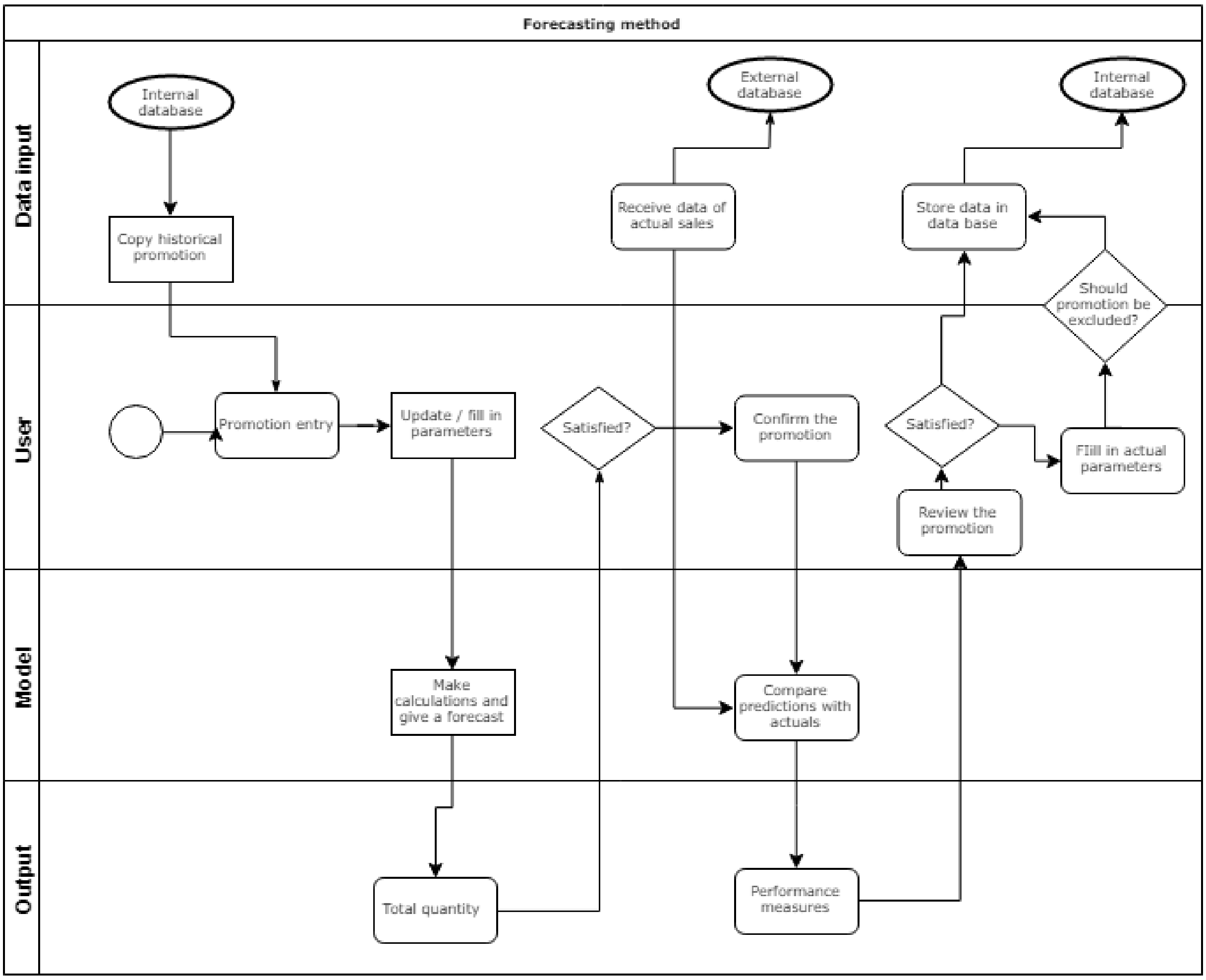
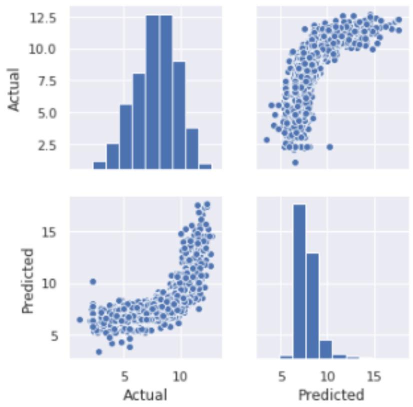
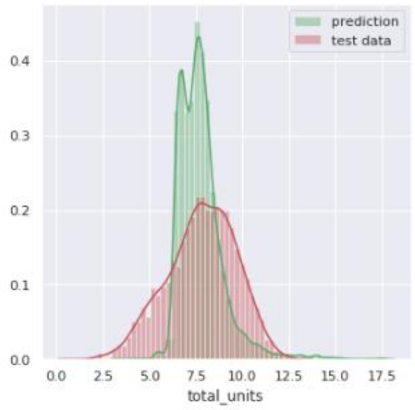


Figure 35: Flowchart of the realized forecasting method

Appendix G – LASSO data analysis

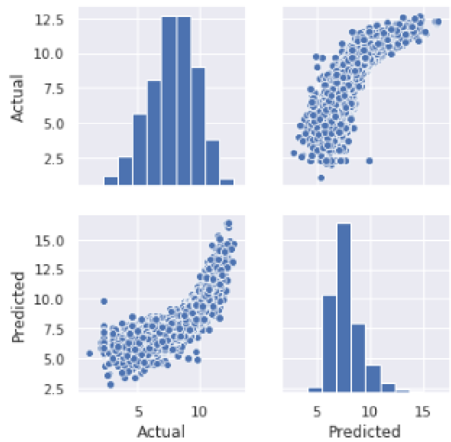
Results using alpha = 10

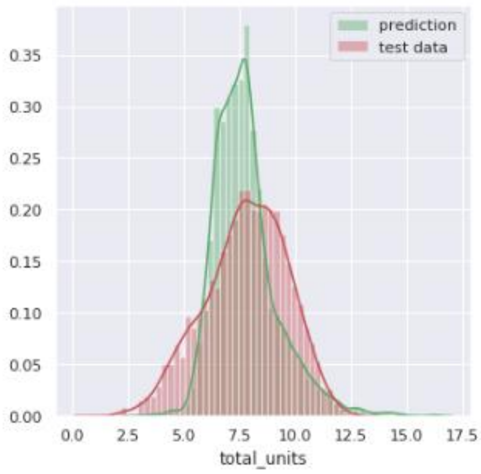





```
Compared to LN_total_units:
R²:-0.0744811410212265
MAE:1.0174845955981477
MAPE:15.5293463724
Forecast Error: 13.033759523659151
```

Results using alpha = 1

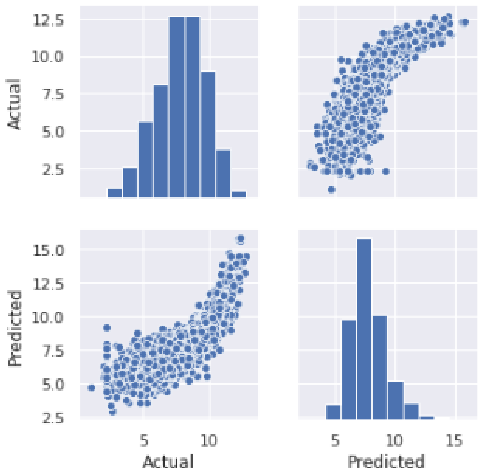


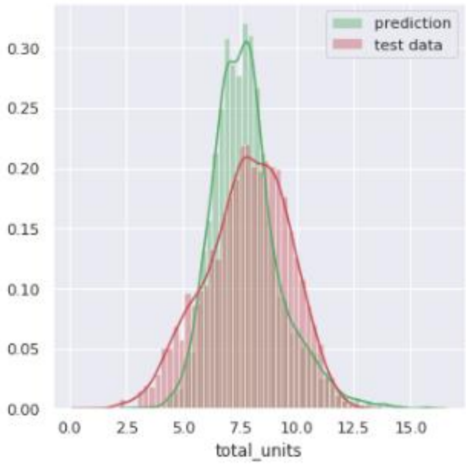



```
Compared to LN_total_units:
R²:0.4236533624714808
MAE:0.8601291792271255
MAPE:13.2864263505
Forecast Error: 11.018070376523236

 number of features used: 19
```

Results using alpha = 0,1

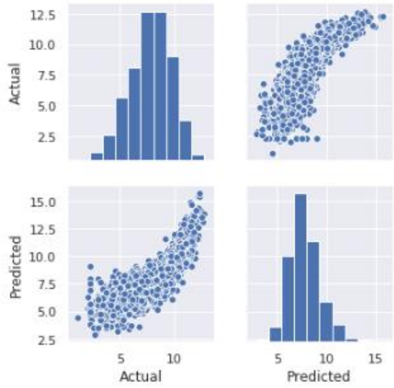


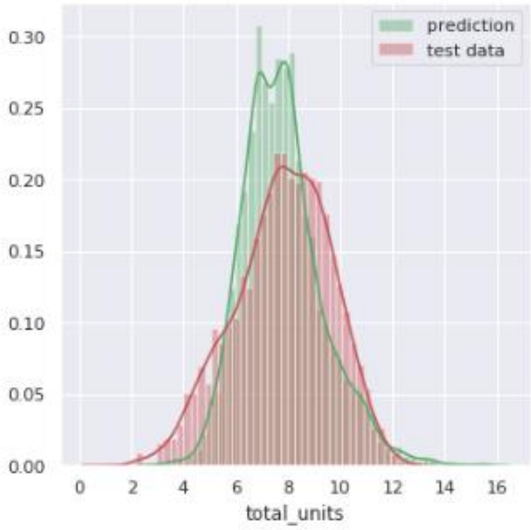



```
Compared to LN_total_units:
R²:0.537008157323154
MAE:0.8114042680473652
MAPE:12.4240823309
Forecast Error: 10.393914710800052

 number of features used: 31
```

Results using alpha = 0,01

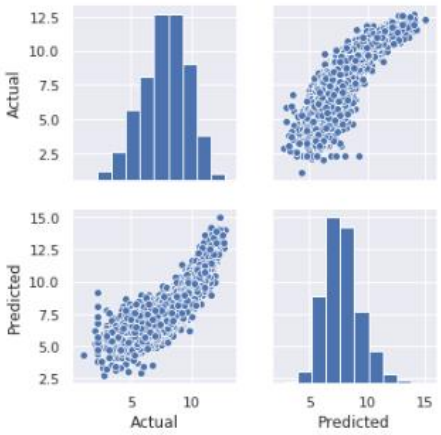


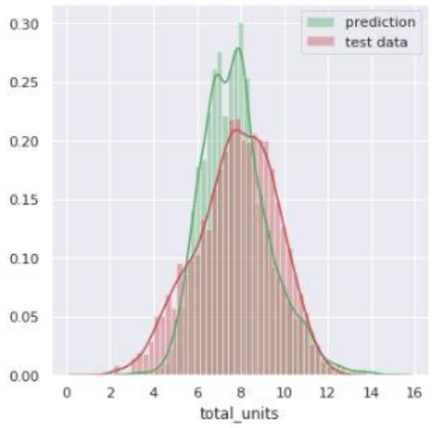



```
Compared to LN_total_units:
R²:0.584689409048196
MAE:0.7829006288001632
MAPE:11.9907179319
Forecast Error: 10.028789203146772

 number of features used: 41
```

Results using alpha 0,001

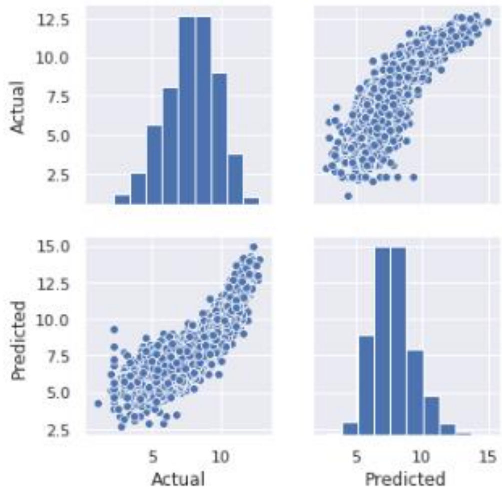


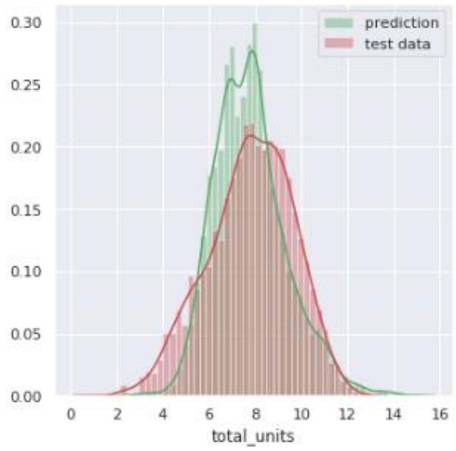



```
Compared to LN_total_units:
R²:0.6154114027475552
MAE:0.7615275515400146
MAPE:11.6904759387
Forecast Error: 9.75500466577439

 number of features used: 55
```

Results without using alpha -> using Linear Regression






```
Compared to LN_total_units:
R²:0.6177555519780452
MAE:0.7606557990305919
MAPE:11.6766530463
Forecast Error: 9.743837703030074

 number of features used: 66
```

Appendix H – Linear Regression equation

Regression Equation

ln(total_units) = 7,297 + 0,000110 baseline_units - 0,19593 base_price
- 0,01808 planned_discount_perc + 0,03773 previous_promotion_week_distanc
+ 0,033271 second_placement_perc + 0,0 second_placement_yn_0
- 1,0897 second_placement_yn_1 + 0,4460 mechanism_type_0
+ 0,0 mechanism_type_1 + 0,354 mechanism_type_2 + 0,2059 mechanism_type_3
+ 0,2860 mechanism_type_4 + 1,017 mechanism_type_5 + 0,1286 discount_amt
+ 0,04670 discount_perc + 0,2081 multi_buy_y + 0,01377 planned_base_price
- 0,01880 planned_promoted_price + 0,00811 product_volume_per_sku
+ 0,000001 total_baseline_units - 0,003315 total_nr_products
+ 0,5270 product_dimension_13_0 + 0,0 product_dimension_13_1
- 0,1325 product_dimension_13_2 - 1,0412 product_dimension_13_3
+ 0,2670 product_dimension_13_4 + 0,0 product_dimension_17_0
- 2,02 product_dimension_17_1 - 0,59 product_dimension_17_2
- 1,25 product_dimension_17_3 - 1,15 product_dimension_17_4
- 1,13 product_dimension_17_5 + 0,30 product_dimension_17_6
- 0,66 product_dimension_17_7 - 0,54 product_dimension_17_8
- 1,06 product_dimension_17_9 - 1,55 product_dimension_17_10
- 1,48 product_dimension_17_11 - 1,16 product_dimension_17_12
- 0,08 product_dimension_17_13 - 1,79 product_dimension_17_14
- 0,50 product_dimension_17_15 - 1,71 product_dimension_17_16
- 1,81 product_dimension_17_17 - 1,11 product_dimension_17_18
- 1,90 product_dimension_17_19 - 1,42 product_dimension_17_20
- 0,91 product_dimension_17_21 - 0,14 product_dimension_17_22
- 1,43 product_dimension_17_23 - 0,94 product_dimension_17_24
- 0,91 product_dimension_17_25 - 2,66 product_dimension_17_26
- 0,89 product_dimension_17_27 - 1,23 product_dimension_17_28
- 0,73 product_dimension_17_29 - 0,05 product_dimension_17_30
- 0,28 product_dimension_17_31 - 0,97 product_dimension_17_32
- 1,73 product_dimension_17_33 - 0,25 product_dimension_17_34
- 0,91 product_dimension_17_35 - 0,47 product_dimension_17_36
- 3,81 product_dimension_17_37 - 3,12 product_dimension_17_38
- 0,983 product_dimension_17_39 - 1,260 product_dimension_17_40
+ 0,0 discount_perc_cohort_0 + 0,1364 discount_perc_cohort_1
+ 0,0355 discount_perc_cohort_2 - 0,3594 discount_perc_cohort_3
+ 0,2288 discount_perc_cohort_4 - 0,970 discount_perc_cohort_5
- 1,987 discount_perc_cohort_6 + 0,0 planned_discount_perc_cohort_0
+ 0,4152 planned_discount_perc_cohort_1
+ 0,2763 planned_discount_perc_cohort_2
+ 0,6577 planned_discount_perc_cohort_3
+ 0,477 planned_discount_perc_cohort_4
+ 0,546 planned_discount_perc_cohort_5
+ 1,172 planned_discount_perc_cohort_6 + 0,0 week_1 + 0,1911 week_2
+ 0,0217 week_3 + 0,1115 week_4 + 0,0810 week_5 + 0,1384 week_6
+ 0,0933 week_7 + 0,1659 week_8 - 0,0406 week_9 + 0,1134 week_10
- 0,0126 week_11 + 0,1213 week_12 + 0,1762 week_13 - 0,0095 week_14
+ 0,1176 week_15 + 0,1335 week_16 + 0,1697 week_17 + 0,0954 week_18
- 0,0564 week_19 + 0,1790 week_20 + 0,2485 week_21 + 0,2108 week_22
+ 0,0271 week_23 + 0,1430 week_24 + 0,2495 week_25 + 0,0386 week_26
+ 0,0741 week_27 + 0,083 week_28 + 0,1992 week_29 + 0,0419 week_30
+ 0,0381 week_31 - 0,0146 week_32 + 0,0942 week_33 + 0,1515 week_34
+ 0,1116 week_35 + 0,0591 week_36 + 0,1808 week_37 + 0,0405 week_38
+ 0,0975 week_39 + 0,2620 week_40 - 0,1108 week_41 + 0,0314 week_42
+ 0,0407 week_43 + 0,2848 week_44 + 0,3488 week_45 + 0,0977 week_46
+ 0,1285 week_47 + 0,3330 week_48 - 0,1388 week_49 + 0,0649 week_50
+ 0,1658 week_51 - 0,219 week_52

Appendix I – List of independent variables in each model

This section describes the list of independent variables in each different LASSO model (see Table 21).

Table 21: The list of independent variables included in each model

Linear Regression	Lasso0.0001	Lasso0.001	Lasso0.01	Lasso0.1	Lasso	Lasso10
account_banner						
account_id						
base_price	x	x	x	x		
baseline_units	x	x	x	x	x	x
baseline_units_e1t	x	x	x	x	x	x
baseline_units_int	x	x	x	x	x	x
baseline_vol	x	x	x	x	x	x
consumer_length	x					
discount_amt						
discount_perc	x	x	x	x	x	
e1clude_yn						
field_10002_8	x	x				
field_14052						
field_14058	x	x				
mechanism	x	x	x	x		
mechanism_type	x	x	x			
multi_buy_1	x	x				
multi_buy_y	x	x				
original_pid	x	x	x	x	x	x
pid	x	x	x	x	x	x
planned_base_price	x					
planned_discount_amt		x				
planned_discount_perc	x	x	x	x		
planned_promoted_price	x	x	x			
previous_promotion_week_distance	x	x	x	x		
prod_desc	x	x	x	x	x	
product_alt_uom_per_cu	x	x				
product_cu_per_sku	x	x	x	x		
product_dimension_13	x	x	x			
product_dimension_14	x	x	x	x		
product_dimension_17	x	x	x			
product_dimension_190	x	x				
product_dimension_194	x	x	x	x	x	
product_dimension_20	x	x	x	x	x	
product_volume_per_sku	x	x	x	x		
promoted_price	x	x	x			
promotion_dimension_13	x	x				
promotion_dimension_14	x	x	x	x	x	
promotion_dimension_17	x	x	x			
promotion_dimension_190	x	x	x	x		
promotion_dimension_194	x	x	x	x	x	
promotion_dimension_20	x	x	x	x	x	x

promotion_dimension_4	x	x	x				
promotion_e1t_id	x	x	x	x	x	x	
promotion_name	x	x	x	x			
promotion_status	x	x	x	x			
promotion_type	x	x	x				
relative_start_week	x						
second_placement_perc	x	x	x	x	x	x	
second_placement_yn	x	x	x				
template	x	x	x	x			
week	x	x	x	x			
yearweek	x	x					
total_baseline_units	x	x	x	x	x	x	
total_baseline_vol	x	x	x	x	x	x	
total_nr_products	x	x	x	x	x	x	
total_fwb_perc	x	x	x	x	x	x	
discount_perc_cohort	x	x	x				
planned_discount_perc_cohort	x	x					
field_10002							
field_10002_11	x	x					
field_10002_13	x	x					
field_10002_16	x	x					
field_10002_12	x						
field_10002_4	x						
field_10002_5	x	x					
Total number of variables:	**66**	**59**	**55**	**41**	**31**	**19**	**13**

Appendix J – Ordinary Least Square regression results

This section present the results of the ordinary least square methods (see Figure 36)

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            total_units   R-squared:                       0.706
Model:                            OLS   Adj. R-squared:                  0.705
Method:                 Least Squares   F-statistic:                     2061.
Date:                Mon, 16 Sep 2019   Prob (F-statistic):               0.00
Time:                        18:29:53   Log-Likelihood:            -3.8067e+05
No. Observations:               35275   AIC:                         7.614e+05
Df Residuals:                   35233   BIC:                         7.618e+05
Df Model:                          41
Covariance Type:            nonrobust
====================================================================================================
                                       coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------------
const                            -1.245e+04   3229.443     -3.854      0.000   -1.88e+04   -6117.397
base_price                       -3009.4877    145.399    -20.698      0.000   -3294.474   -2724.501
baseline_units                       3.2872      0.043     75.844      0.000       3.202       3.372
baseline_units_ext                   0.1537      0.029      5.379      0.000       0.098       0.210
baseline_units_int                  -0.1069      0.031     -3.478      0.001      -0.167      -0.047
baseline_vol                         0.9240      0.036     25.863      0.000       0.854       0.994
discount_perc                      150.4563     25.183      5.975      0.000     101.097     199.815
mechanism                          -48.1423     10.653     -4.519      0.000     -69.023     -27.261
mechanism_type                     521.5741    164.447      3.172      0.002     199.253     843.895
original_pid                         0.5190      0.221      2.344      0.019       0.085       0.953
pid                                 -0.2591      0.156     -1.662      0.097      -0.565       0.047
planned_discount_perc              123.3467      8.284     14.890      0.000     107.110     139.583
planned_promoted_price              72.2932     13.350      5.415      0.000      46.127      98.460
previous_promotion_week_distance    36.1266     50.552      0.715      0.475     -62.956     135.210
prod_desc                           -3.1153      0.414     -7.516      0.000      -3.928      -2.303
product_cu_per_sku                   3.0745     19.086      0.161      0.872     -34.334      40.483
product_dimension_13              1189.7601    137.224      8.670      0.000     920.796    1458.724
product_dimension_14               -15.0019     17.305     -0.867      0.386     -48.920      18.916
product_dimension_17               -12.3793     77.918     -0.159      0.874    -165.100     140.342
product_dimension_194                1.1545      1.918      0.602      0.547      -2.605       4.914
product_dimension_20                 8.8131      1.112      7.922      0.000       6.633      10.994
product_volume_per_sku             -66.8898     26.313     -2.542      0.011    -118.465     -15.315
promoted_price                    4211.1145    219.012     19.228      0.000    3781.843    4640.386
promotion_dimension_14             -47.3047      7.031     -6.728      0.000     -61.085     -33.524
promotion_dimension_17             633.2579    119.255      5.310      0.000     399.514     867.002
promotion_dimension_190           -474.2459    134.775     -3.519      0.000    -738.409    -210.083
promotion_dimension_194             -7.9957      1.491     -5.364      0.000     -10.917      -5.074
promotion_dimension_20               4.9515      0.670      7.392      0.000       3.639       6.265
promotion_dimension_4                8.6201      2.745      3.141      0.002       3.240      14.000
promotion_ext_id                    -0.0476      0.013     -3.601      0.000      -0.073      -0.022
promotion_name                       1.8261      0.941      1.941      0.052      -0.018       3.670
promotion_status                    27.1392      5.398      5.028      0.000      16.560      37.719
promotion_type                    -259.2259     46.721     -5.548      0.000    -350.800    -167.652
second_placement_perc              148.1931      4.528     32.726      0.000     139.318     157.069
second_placement_yn               8255.4240    399.641     20.657      0.000    7472.116    9038.732
template                            19.7501     28.995      0.681      0.496     -37.080      76.580
week                               -18.2580      4.586     -3.981      0.000     -27.247      -9.269
total_baseline_units                -0.0008      0.001     -0.723      0.469      -0.003       0.001
total_baseline_vol                  -0.0452      0.002    -21.143      0.000      -0.049      -0.041
total_nr_products                   -3.8008      1.288     -2.950      0.003      -6.326      -1.276
total_fwb_perc                      -0.1948      0.134     -1.454      0.146      -0.457       0.068
discount_perc_cohort              1904.4455    224.466      8.484      0.000    1464.485    2344.405
==============================================================================
Omnibus:                    54977.243   Durbin-Watson:                   1.613
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         81957031.413
Skew:                           9.558   Prob(JB):                         0.00
Kurtosis:                     238.363   Cond. No.                     9.70e+06
==============================================================================
```

Figure 36: Output of the ordinary least square regression method using Python.

Appendix K – Model results of comparable research

This section describes the model results of comparable research from literature. Over the years, multiple forecasting models are developed that aim to forecast shopper demand. The model results of comparable research are obtained from Derks (2015) and are shown in Table 22.

Table 22: Result from Derks (2015) on the performance of comparable forecasting models found in literature.

	Derks	Van Donselaar et. al	Peters	Van der Poel	Van den Heuvel	Van Loo	Cooper et. al
Year	2015	2015	2012	2010	2009	2006	1999
Point of view	Manufacturer	Retailer	Retailer	Manufactuer	Retailer	Retailer	Retailer
Aggregation level	Supply Chain	Supply Chain	Supply Chain	Supply Chain	Supply Chain	Supply Chain*	Store
Method	Prais Winsten regression	Lineair + quadratic model based on category	Ordinary least squares	Ordinary least squares	Ordinary least squares	Ordinary least squares	Ordinary least squares
Dependent variable	LN Lift factor	LN Lift factor	LN Lift factor	LN Lift factor	Lift factor	P(LF)	LN_Sales
Sample size	3611	N.a.	2175	1238	N.a.	1556	N.a.
# Predictors	34	24	21	21	8	39	67
Adj. R2	0.742	N.a.	0.767	0.691	0.44	0.45	N.a
MAPE validation (best)	26.68%***	21.00%	28.18%	31.30%	N.a.	31.14%	N.a
* Best model is based on Supply Chain level	*** The MAPE of model 1.ai is taken here						
** OLS = Ordinairy Least Squares linear regression							

Appendix L – Confidential

Appendix M – Confidential