August 2020

Workforce Prediction of Order Picking Personnel

Predicting the Required Number of Flex Order Picking Personnel in Distribution Centres of Albert Heijn

PUBLIC VERSION

Master Thesis by:

N.J. Huitink

S1470701

Supervisors University of Twente:

dr. ir. M.R.K. Mes

dr. E. Topan

Supervisor Albert Heijn:

P. Meints Msc.

Management Summary

Workforce management is a complex task consisting of multiple stages. The goal is to match workload with workforce, by consecutively predicting the workload, designing shifts, and scheduling the employees. At the distribution centres of Albert Heijn, the shifts are fixed, thus the workforce management process simply consist of determining the workload and assigning the employees to shifts. To fulfil the entire workforce requirements for the order picking tasks in the distribution centres, Albert Heijn employs some employees themselves. However, the largest part of the workforce is fulfilled by flex employees of contracted employment agencies.

The contracted employment agencies allow a *confidential* % in- or decrease in the requested number of flex employees per shift, based on the request made one week in advance. However, this flexibility in the number of employees is taken into account in the tariffs of the agencies. Improving the accuracy of the prediction of the required number of flex employees one week in advance, could reduce the marge and therefore reduce the tariffs of the employment agencies.

The aim of this research is to identify possibilities to improve the prediction accuracy of the required number of flex order picking employees. To do so, both the current prediction method is analysed for improvements, and four data mining models are tested. The models are based and tested on the data of the distribution centre in Zwolle and focus on the non-perishable pick zone. The data of 2019 and the beginning of 2020 is used as training and test data.

The four different data mining techniques that have been applied to predict the required number of flex employees are: Generalized Linear Models, Deep Learning, Gradient Boosted Trees, and Random Forest.

Improving the Current Prediction Method

The current prediction method uses a simple calculation to determine the required number of flex employees based on four uncertain variables, as shown in the equation below.

$$Flex\ Employee\ Hours = \frac{Demand * Service\ Percentage}{Productivity} - (Planned\ AH\ Hours + AH\ Illness)$$

It is possible that not all items ordered by the stores can be picked, since they might not be available at the distribution centres. The percentage of orders that can be fulfilled is called the service percentage. Thus, the total number of ordered products, times the service percentage is the number of colli to be picked. Both those values are uncertain and thus predicted or estimated. Based on the predicted number of colli to pick and the predicted order picking productivity, the expected order picking hours can be determined by simply dividing. Reducing those hours with the expected hours fulfilled by AH employees, the expected number of flex hours is obtained. The only uncertainty in the available AH hours, is that AH employees might call in sick last minute.

Improving the Colli Prediction

Since the numerator is the most uncertain (the total number of colli to pick), the four machine learning methods are tested to directly predict the number of colli to pick. The Generalized Linear Model was the best of the four models and was able to outperform the current method in terms of predicting the number of colli to pick. However, the resulting number of predicted flex employees was less accurate. This indicates that the error in the colli prediction, is somehow accounted for by the errors in the expected productivity or AH hours. Either way, this method does not seem promising and it is therefore not advised to use machine learning to alter the colli predictions.

Directly Predicting the Number of Flex Employees

Directly predicting the required number of flex employees did show promising results. The best performing model was the Generalized Linear Model (GLM) for which the most important features where, the shift and day of week, the predictions given by the replenishment department, and the average realizations of the previous three weeks.

By using the GLM, the MAPE of the prediction could be reduced from *confidential* % to *confidential* %, as shown in the table below. Furthermore, the percentage of time the actual required number of flex employees is within a *confidential* % range of the predicted value was increased from *confidential* % to *confidential* %. This improves the position of Albert Heijn in the negotiations with the employment agencies. Since those results are only based on the regular colli pick zone of the non-perishable department, the performance could improve even further in case the other departments are included as well. This is the case, since the total number of employees increases, also making the *confidential* % range larger.

<table>
<tr><th></th><th>Current</th><th>Predict Flex Directly</th></tr>
<tr><td>MAD</td><td colspan="2" rowspan="3">confidential</td></tr>
<tr><td>MAPE</td></tr>
<tr><td>% in confidential %</td></tr>
</table>

This GLM model was also analysed in more detail. With five experiments it is shown that it is possible to improve the performance of the model in case under- or overestimation is expected to be more costly. This is done by altering the predicted value to another value within the given prediction interval.

Based on the promising results, it is recommended to Albert Heijn to continue the research into the usage of Generalized Linear Models to predict the required number of flex employees. In addition to the improved predictions, using this method also reduces a lot of manual work. Since each distribution centre has their own capacity planner, the automation reduces the required work of five employees. Additionally, the automation reduces the possibility of human errors in the calculations.

Preface

On your screen, you have the final version of my master thesis. Although most of you would have read my thesis online anyways, the current situation has most likely forced everyone to read everything online. This situation has required, and continues to require, changes and flexibility from everyone. It has also made finishing up my master stranger than I expected. However, with the help of the right people during my thesis and in this strange end phase, I am glad to hand in this final project of my masters. I would like to use this preface to thank all those people.

First of all, I would like to thank Martijn Mes and Engin Topan from the University of Twente for their supervision. Based on your input and feedback, I was challenged to broaden my scope and get acquainted with subjects outside the curriculum I had followed thus far. Additionally, I would like to thank Cornelis ten Napel, from the University of Twente as well, who helped me with great personal advice to handle the final challenges.

Off course I would also like to thank my supervisor from Albert Heijn, Pieter Meints. With your energy and passion for Albert Heijn, you made it easy to enjoy working at Albert Heijn. You really made sure I felt as a part of the team and that I had everything that I required to work on my thesis. I would also like to thank all my other colleagues at the Logistics Preparation department at Albert Heijn for their openness and support. Additionally, I would like to thank all other Albert Heijn employees who have contributed to this project. From site managers, to headquarter employees, HR employees, and operational personnel at the distribution centers, you have all been very helpful.

Last, but definitely not least, I would like to thank my friends and family who have all helped me through the hardest times. Especially Joris, I could not have done this without you.

I hope you enjoy reading this thesis.

Nienke Huitink

Utrecht, August 2020

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Description
DC(s)	Distribution Centre(s)
GLM	Generalized Linear Model
LDC	*Dutch: Landelijk Distributie Centrum* English: National Distribution Centre Processes “slow moving” products which are delivered to RDCs, where they are cross-docked and delivered to the stores.
MAPE	Mean Absolute Percentual Deviation
RDC	*Dutch: Regionaal Distributie Centrum* English: Regional Distribution Centre Processes “fast moving” products which are directly delivered to the stores.
WAB	*Dutch: Wet Arbeidsmarkt in Balans* English: Balanced Labour Market Act
WFM	Workforce Management
WMS	Workforce Management System

1. Introduction

Within this thesis, the workforce prediction of the number of required order picker personnel in distribution centres of Albert Heijn is addressed. Albert Heijn is a Dutch supermarket chain, which operates grocery stores and smaller "to go" shops, located at, e.g., stations or airports. Additionally, Albert Heijn operates order pickup points and offers a delivery service for groceries. The chain originates from a small family grocery store started in 1887. Since then, the family business has grown into the largest Dutch supermarket chain, with a market share of 34% in 2018 (Meijsen, 2019). Albert Heijn is part of Ahold Delhaize, which is one of the world's largest food retail groups. They are a leader in supermarkets and e-commerce, and a company at the forefront of sustainable retailing (Ahold Delhaize, 2019).

The motivation for this research is given in Section 1.1. The problem at hand is described in more detail by introducing the current workforce management methods at Albert Heijn in Section 1.2. Section 1.3 describes the research aim and demarcation, which are then comprised into research questions to be answered, in Section 1.4, including a description of the structure of the remainder of this report.

1.1. Research Motivation

Employees with a permanent contract often have better working conditions and more rights than flex employees. To reduce this gap between those types of employees, the Dutch government adjusts the Balanced Labour Market Act, in Dutch WAB, *"Wet Arbeidsmarkt in Balans"* (Rijksoverheid, 2019). The adjustment is put in place at the first of January 2020. This update in the Dutch law is the main motivation for this research.

The most important update in the WAB related to workforce management at the distribution centres, is the regulation concerning the call period. From the first of January 2020, the employer must inform the employee at least four days in advance of the required working hours. This information must be given in written notification or electronically. In case the employer requests the employee later than four days in advance, the employee is not obligated to show up. On the other hand, if the employer cancels the promised working hours less than four days in advance, the employee is entitled to the payment of the promised working hours (Rijksoverheid, 2019).

Within the distribution centres of Albert Heijn, almost *confidential* % of the required operational workforce is fulfilled by flex workers, who mainly perform the order picking tasks. Those flex workers ensure flexibility within the distribution centres since their working hours can easily be adjusted. However, the update of the WAB reduces this flexibility, either resulting in a decrease of the flexibility of the distribution centres or an increase in the costs to operate the distribution centres with the same flexibility.

The first estimates made by Albert Heijn indicated an expected increase of *confidential* euros per year to preserve the current flexibility level within the operational workforce of the distribution centres. This estimation is based on the prices the employment agencies asks for their service of offering the flex employees. The employment agencies are responsible to pay their employees, even in case Albert Heijn makes last minute adjustments to the schedules of the flex employees. Since Albert Heijn requires this flexibility from the employment agencies, the agencies simply cope with those costs by raising their rates for Albert Heijn.

However, while conducting this research, the contracts with the employment agencies have been renewed. In this process, multiple employment agencies can make offers concerning their conditions and tariffs. Since multiple employment agencies are interested in working for Albert Heijn, they all make their best bids. As a result of good market forces, the updated contracts did not increase in prices due to the new four day calling period rule. However, that does not mean that the next time the contracts are updated, the market forces are the same and result in those low tariffs again.

All contracts with the employment agencies are specified such that Albert Heijn is allowed to up- or downscale the request for the number of flex employees with *confidential* %, based on the request made one week in advance. This flexibility is taken into account by the employment agencies in their rates. It is known that the employment agencies are able to reduce their tariffs in case the up- and downscaling rule of *confidential* % could be reduced. Although the updated four-day rule did not affect the costs at Albert Heijn, improving the prediction accuracy will still be beneficial when the contracts with the employment agencies are updated again.

Although the flexibility is costly and might become more expensive when the contracts with employment agencies are updated again, internal changes within Albert Heijn require even more flexibility of the distribution centres. For example, the rapid growth of AH Online, allowing customers to order online. Those orders are harder to predict and can vary last minute. This increases the uncertainty in demand volumes, resulting in more uncertainty in workforce requirements for the order pickers within distribution centres.

Research Focus and Goal

As will be explained in more detail in the literature review in Section 2.1, workforce management consists of four phases: workload prediction, staffing, shift scheduling, and rostering. Based on the workload as predicted in the first stage, in the staffing phase the total number of employees in the employee pool should be determined. Furthermore, the predicted workload is input in determining the required shifts to fulfil the workload. Finally, once shifts are determined, the employees should be rostered into the shift such that the rosters of each employee comply with the rules and regulations, and all shifts are covered.

The workload prediction phase of workforce management is the focus of this research. Due to the focus on the flex employees, the staffing phase is irrelevant, since the employment agencies are responsible to ensure sufficient staff levels. Furthermore, the order picking shifts at the distribution centres are fixed. Finally, the rostering task of the flex employees is again performed by the contracted employment agencies.

Thus, in case of the flex order pickers at the distribution centres, the workload predictions are the most important stage of the workforce management for Albert Heijn. The workload predictions are made per shift and are thus directly translated into the forecasted number of required employees, or the workforce prediction.

This workforce prediction of the flex order pickers is a complex task. Each distribution centre of Albert Heijn has its own way of working, as a result of trial and error in the previous years. However, each distribution centre still faces the same challenges in incorporating uncertainty in their methods. For example, it is known that the productivity, or speed of working of employees, can vary significantly. Furthermore, the exact amount of orders to pick on a certain day can deviate from the predicted value. Finally, full-time AH employees might unexpectedly be absent for example due to illness, increasing the need for flex workers.

All the uncertainties influence the total required number of flex employees during a production day. The exact values only become known during the actual production day itself. However, the number of flex workers and their working hours should ideally be fixed four days in advance, limiting the lastminute costs due to the WAB update.

This research aims to predict the required number of flex employees accurately, based on the information that is known at least one week in advance, to comply with the agreements with the employment agencies and the four-day calling period update in the WAB. To do so, the research is focussed on the uncertain variables affecting the required number of employees and possibilities to improve the current methods.

1.2. Operational Workforce in Distribution Centres of Albert Heijn

Within this section, a brief introduction is given concerning the distribution centres of Albert Heijn and their operational workforce management. A more detailed description is given in Chapter 3, the context analysis. However, to get a better understanding of the problem, the basics and some details are described within this section.

Distribution Centres

In total, Albert Heijn uses eleven distribution centres, from now on abbreviated to DCs. The operations of six of those DCs are outsourced to external warehouse management companies. However, the other five DCs are managed and operated by Albert Heijn employees. The distribution of the DCs throughout the Netherlands is shown in Figure 1.

In Figure 1, the DCs that are managed by Albert Heijn are marked with the blue logo. The distribution centre in Geldermalsen contains slow moving products, which are distributed to other DCs where they are cross docked to the stores. This DC is therefore called a national distribution centre, in Dutch "Landelijk Distributie Centrum" (LDC). The DCs in Zwolle, Zaandam, Pijnacker, and Tilburg contain fast moving products. The products from those DCs are combined with the cross docked products and then directly delivered to the stores. Those DCs are therefore called regional distribution centres (RDC). This process is also depicted in Figure 2, which shows the supply chain of Albert Heijn. The figure also shows returns from stores back to the DCS and suppliers. The returns do not require order picking flex employees and are thus left out of this research.

The DCs managed externally, marked with the black square in Figure 1, are either freezer warehouses, which deliver directly to the stores, or warehouses that contain slow moving products that are cross docked to stores through an RDC, similar to the products as delivered by the LDC.

Figure 1: DCs of Albert Heijn (Meints, 2017)

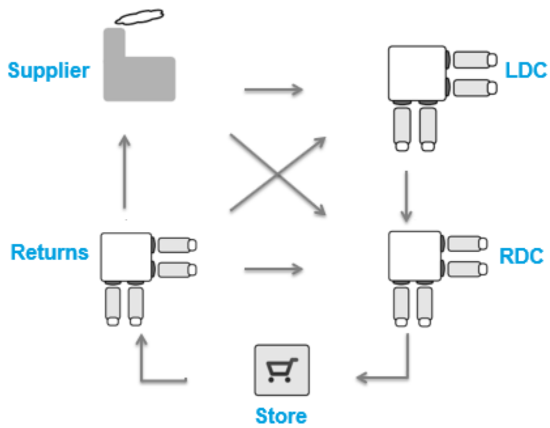

Figure 2: Overview AH Supply Chain (Meints, 2017)

The Operational Workforce

The operational workforce within the DCs of Albert Heijn is defined as the employees that perform the daily tasks in the DCs. This roughly consist of four types of tasks: the order picking, truck loading and unloading, forklift operations, and other smaller tasks such as cleaning and counting stock. All those operational employees are required to facilitate the basic product flow as depicted in Figure 3.

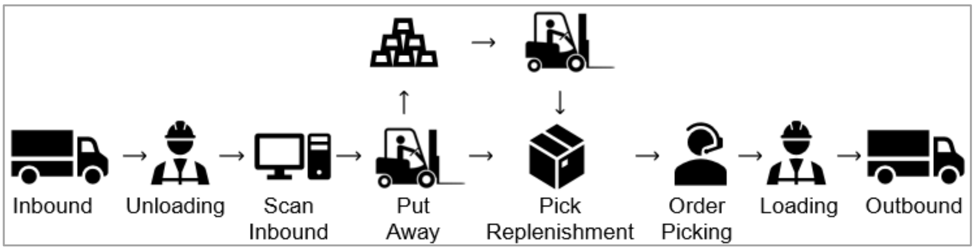

Figure 3: Basic Product Flow through the Distribution Centre

Truck drivers delivering goods at the distribution centre must notify their arrival. The truck is then assigned to a dock where the products are unloaded. To handle the inbound truck, two types of personnel are required: unloaders and forklift drivers. The truck driver is responsible to unload the truck together with an unloading employee of the DC. This employee is also responsible to check the products that are delivered. Each container must be scanned, to process the arrival within the Warehousing Management System (WMS).

Once the unloading and scanning is finished, forklift drivers are required to move the products from the inbound lane to the assigned storage location. This type of forklift movement is called "put away". This put away is either to a pick location or a buffer location. At a pick location, the product is stored in such a manner that an order picker is able to perform a pick. A buffer location is one from which a pick is not possible, for example a higher place in a storage rack. Products are stored in those locations and are later retrieved to replenish a pick location. This replenishment is again performed by a forklift driver and is called "pick replenishment". In general, a forklift driver is either assigned to the put away task, or the pick replenishment task. However, no special skill or knowledge is required to perform either of those tasks, thus the personnel is in theory interchangeable.

The order picking task is called "production", since the pickers produce the output of the distribution centre. This task takes up most of the operational workforce (in hours) in the DCs, namely at least *confidential* % of the total hours. The second most frequent task are the forklift movements, those require only at most *confidential* % of the total hours. This can be seen in Table 1 below.

Table 1: Hours Spent Per Task per DC, % of Total, Week 47, 2019

Task	Zaandam	Tilburg	Zwolle	Pijnacker	Geldermalsen	Total
Production			*Confidential*			
Forklift Movements						
Loading						

The order pickers in the DCs of Albert Heijn use the "voice picking" method, by which they receive and confirm order picking tasks through headsets. The order pickers of most non-perishable products drive a cart, carrying five roll containers that must be filled with products. However, the order pickers of most perishable products fill one container at a time, loading the container with crates containing the products. The route the pickers must travel and the number of products of a certain type they must pick, is given through their headsets. Once all products are picked, the order picker must stage the roll containers on the correct outbound lane, which is also instructed through the headset. The order picker is then assigned to a new order. This is not a randomly selected order to pick, it is dependent on the trucks that are scheduled to leave the DC in the coming period.

Once all orders of a shipment are picked and staged at the correct outbound staging lane, those containers can be loaded into the truck. This is done by the truck driver and an employee of the distribution centre. Each truck is assigned to a specific time frame, which will make sure the truck is able to deliver the products at the grocery stores in time.

In addition to the basic product flow, the RDCs also perform cross-docking activities. At Albert Heijn this is called the "transito flow". As shown in Figure 2, this type of product flow starts in a national distribution centre, an LDC. Trucks arriving from the LDC typically contain roll carts for several distinct outbound trucks. The roll carts are temporarily staged at the "transito lanes", from which they are obtained once the outbound truck is ready to be staged. Cross-docking employees are required to fetch the roll carts.

Finally, to keep the distribution centres clean and easy to work in, personnel is required to clean the isles. A forklift driver is required to move empty pallets. Furthermore, an employee with a cart containing trash cans and clean-up material is responsible to clean the pick locations, removing empty cartons etcetera. This employee might also be required in case accidents happen, causing spillage of goods.

During a production day, all operational personnel is managed by the cockpit employees. The cockpit is an office in the centre of the distribution centre, where often three employees are active. The first is responsible to manage the inbound trucks by receiving their arrival notification and assigning them to their docks. The second manages the outbound process. Finally, the third manages all other operational personnel. For example, re-assigning order pickers to a forklift assignment in case a lot of pick replenishment is required. Furthermore, the cockpit employees are responsible to monitor the progress of the order picking process, by assessing whether all orders will be picked in time, or if the orders will be picked too fast. In the latter case, the cockpit employee can decide to send some order pickers home early. In case the order pickers seem to run out of time, additional employees might be called in, or orders are cancelled.

Workforce Planning

In addition to the operational management performed by cockpit employees, each DC also has a capacity planner who is responsible for the main part of the workforce management process. The capacity planner is responsible to determine the daily required workforce of each employee type. Some numbers are fixed, for example the number of required cleaning employees. For the other tasks, the capacity planner determines this number based on the expected amount of work and the division of work throughout the production day.

The expected amount of work is defined as the total number of colli to pick from the DC. A colli is defined as one order picking unit of a certain product. An order typically consists of multiple different colli, from which one or more should be picked. For example, an order could consist of ten colli, of which five are boxes containing yoghurt, and five are boxes containing custard. The boxes of yoghurt and custard might contain multiple packages of yoghurt and custard, for example, six per box. See Figure 4 for an illustration of this example.

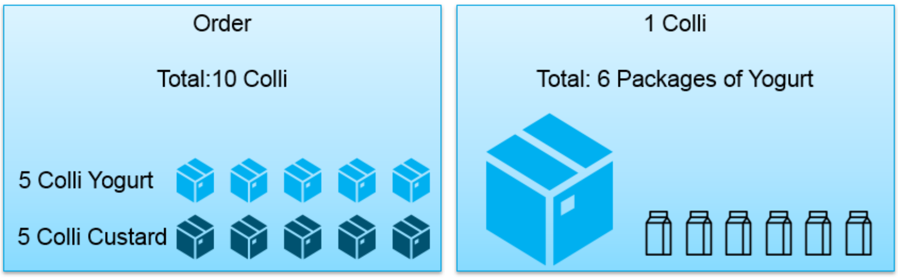

Figure 4: Example Order and Colli

The expected amount of colli to pick is defined per shift, which is either the day, night or evening shift. Not all the DCs use evening shifts, and when those shifts are used, this is often not the case for each day of the week.

To fulfil the workload for a typical day without evening shift, a production day typically consists of three shifts for employees, as shown in Figure 5. The figure shows the shifts as used in Zwolle, but other DCs have similar schedules. The first shift starts at 11 p.m. the day in advance. The second shift at 7 a.m. and the third shift at 8:30 a.m. The last shift ends at 05:00 p.m., which is thus the time that all orders should be picked.

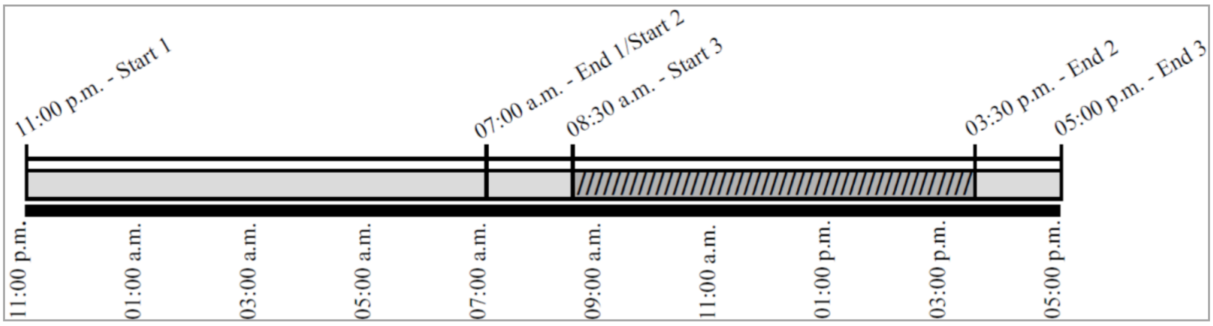

Figure 5: Shift Pattern at DCO (de Lang, 2019)

The expected amount of colli to pick per shift, is determined by the Replenishment department. They make a prediction seven weeks in advance, two weeks and one week in advance of the production day. Based on those colli predictions of the Replenishment department, the capacity planner makes a prediction of the required number of flex employees to schedule.

The data seven weeks in advance is used to inform the employment agency with a rough prediction. Based on the updated colli predictions two weeks and one week in advance of the production day, the capacity planner adjusts the requested number of flex employees. However, the actual number of colli to pick only becomes known after 11 p.m., once the first shift has already started. Based on this information, the capacity planner can adjust the requested number for the day shift. However, as described in the previous section, the requested number of flex employees one week in advance, is limiting the final request on the actual production day. Only up- or downscaling the request made one week in advance with *confidential* % is allowed within the current contracts with the employment agencies.

Although the orders become fixed after 11 p.m., the total number of colli that are actually picked during a shift can still decrease. It might be the case that products are not on stock in the distribution centre. In that case, the products or colli are simply left out of the orders. The percentage of available products in the distribution centre is called the service percentage. The service percentage depends on the delivery of goods from suppliers and the speed of unloading the trucks and replenishing the pick locations. The service percentages are estimated by the capacity planner, to determine the actual number of colli that will be picked during a shift.

The uncertainty in the total amount of colli to pick is not the only uncertainty affecting the number of required order pickers. The productivity of the order pickers, which is the number of colli an order picker picks per hour, can also significantly differ from day to day. For example, when it is very cold, employees tend to work faster than when it is very hot. Or when a lot of order pickers are working at the same time, they might reduce their productivity due to congestions. In case the productivity is lower than expected, there is a change that too few flex employees are requested. On the other hand, if the employees achieve a higher productivity than expected, the employees might run out of work early.

Finally, the required number of flex employees is dependent on the number of Albert Heijn employees that are working. Based on the scheduled Albert Heijn employees, the number of required flex employees is determined. However, unexpected absence of Albert Heijn employees results in an increase in the demand for flex employees. For example, if an Albert Heijn employee calls in sick, an additional flex employee is requested.

1.3. Research Aim and Demarcation

As briefly described in Section 1.1, the workload prediction process for the required number of flex employees is a complex task. Section 1.2 introduces the workforce management process of the operational employees and the current workload prediction method of the capacity planners. The main problem in determining the required number of flex employees, is the uncertainty of multiple important input variables.

The uncertain input variables that are addressed within this thesis are: the uncertainty in the amount of colli that is ordered, the service percentage, the productivity of employees, and the unexpected absence of Albert Heijn employees due to illness. Those are depicted as the uncertain input variables to the decision model in Figure 6.

As will be shown in the context analysis in Chapter 3, the current decision model is rather simple. All uncertain input variables are point forecasts, resulting in a point forecast made by the decision model. Possible adjustments to this simple decision model to improve the performance are also described in the context analysis.

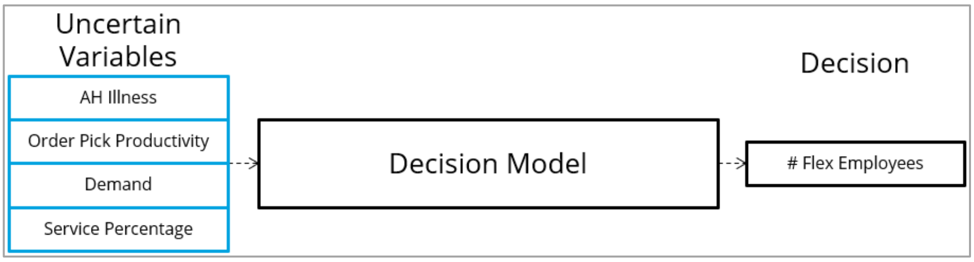

Figure 6: Determination of the Number of Flex Employees in the Current Situation

The aim of this research to make the prediction of the required workforce accurately. This is attempted in two ways. First, by altering the current decision model such that it includes prediction intervals. And second by replacing the current decision model by a machine learning algorithm to predict the required number of flex employees.

It is important that the input, required for both the adjusted version of the current decision model and the machine learning models, is known at least one week in advance, to comply with the update in the WAB and the contracts with the employment agencies.

Demarcation of Data Analysis

Although Albert Heijn stores a lot of data of all DCs, only the DC in Zwolle (Overijssel), DCO, will be analysed within this thesis. Due to time constraints it is not feasible to analyse the data of all DCs. DCO is selected since they collect a lot of relevant additional data, which can be used to enhance the analysis.

Although the scope is narrowed down to DCO, the analysis scopes even further, focussing on the non-perishable pick zone at the DC. This pick zone is the largest, and thus has the largest impact on the total cost of the operational personnel. Some analysis include perishable pick zone A as well, as benchmark and when analysing the possibilities of exploiting economies of scale, as is done in other workforce management problems as described in the literature review in Chapter 2.

1.4. Research Questions and Approach

Based on the aim of the research as described in Section 1.3, the main research objective is formulated as follows.

Research Objective

Improving the prediction of the required number of flex order pickers, based on the information available at least one week in advance.

As described in the previous section, there are at least four uncertain variables currently used as input to predict the required number of flex order pickers. Those variables are the colli demand, service percentage of the DCs, productivity of employees, and unexpected absence of Albert Heijn employees due to illness. Both altering the current prediction method using the uncertain variables, and, applying alternative methods to predict the required number of flex employees, are addressed within this research. The most important constraint when applying alternative methods is the fact that the information used by the model should be available at least one week in advance.

To achieve this research objective, multiple research questions are defined. Those questions are subdivided into four phases: literature review, context analysis, data analysis and mining, and Monte Carlo analysis. In case a research question includes sub-questions, those sub-questions are required to answer the parent research question.

Literature review

The literature review is required to provide the theoretical framework for the research. This framework is required to get familiar with similar problems and existing methods to solve such problems. Thus, the main aim of the literature review is to identify the state of the art concerning uncertainty in workload predictions in distribution centres. However, this is so specific, that a more general literature review is performed, concerning the entire workforce management process, and the applicability to the case at hand is assessed.

1. What is the state of the art concerning uncertainty in workforce management in distribution centres?
    a. What is workforce management?
    b. What types of uncertainty are known in workforce management?
    c. Which solutions are proposed to deal with uncertainty in workforce management?
    d. Which of those methods are applicable to workforce management in distribution centres?

Additionally, a literature review is performed concerning data mining methods that can be used to predict the required number of flex employees.

2. What data mining methods can be used to predict the uncertain variables?
3. What data mining methods can be used to predict the required number of flex employees?

Context Analysis

The context analysis forms the basis for the data analysis. The aim of the context analysis is two-folded. First, it is important to get more in-depth insight in the current practices of Albert Heijn as a basis for the research. Second, during the identification of the current way of working, multiple hypotheses are drawn concerning the uncertainty and the impact of the uncertainty. Those hypotheses are based on the knowledge and experience of the involved employees, or on existing models as identified in the literature review. The hypotheses concern possible improvement opportunities in the current way of working and will help in guiding the following research steps.

4. How does Albert Heijn currently handle the workforce management of order pickers?
    a. How is the required workforce for order pickers predicted?
    b. How is the uncertainty incorporated in the predictions?
    c. How does Albert Heijn cope with deviations between forecasted and realized colli to pick?

Data Analysis and Mining

In the data analysis and data mining phase, the hypotheses drawn in the context analysis will be tested. To do so, the first step of this phase is obtaining and preparing relevant data. In case appropriate data is available the hypotheses can be tested.

5. Which data is available on the uncertain variables?
6. Which data is available on the required number of flex employees?
7. How can the prediction process be improved based on the results of the hypotheses?

Finally, by using the identified data mining techniques from the literature review, the data can be analysed in more detail to discover unanticipated patterns, such as cyclic behaviour or seasonal patterns, or more difficult patterns only able to identify using datamining techniques. By using those techniques, it might be possible to improve the predictions of the uncertain variables to improve the forecast on the required number of flex employees, or directly predicting the required number of flex employees.

8. Can the prediction of the required number of flex employees be improved?
    a. Can the prediction improve by predicting the uncertain variables more accurately?
    b. Can the prediction improve by directly predicting the required number of flex employees more accurately?

The remainder of this thesis follows the structure of the research questions. This is also shown in Figure 7. Chapter 2 answers the research questions based on the literature review. Combining with the context analysis, hypotheses are drawn concerning the workforce management process in Chapter 3. The available data is described in Chapter 4, in which the hypotheses are tested. Based on the results of the hypotheses in the data analysis and the literature review, alternative prediction models are designed in Chapter 5, of which the results are presented in Chapter 6. This report concludes with the conclusions and recommendations in Chapter 7.

Figure 7: Research Approach and Report Structure

2. Literature Review

This chapter concerns the literature review of two main topics. First, uncertainty in workforce management is addressed in Section 2.1. Second, Section 2.2 addresses prediction methods, specifically data mining techniques, which can be applied in the workload prediction process.

2.1. Workforce Management and Uncertainty

Workforce management, as introduced in Section 2.1.1, is the decision on the number of flex employees to hire, is part of this process. Uncertainty is an important factor in the problem at hand and in general in workforce management. Therefore, Section 0 elaborates on the literature on uncertainty in workforce management and the relevance and applicability to the case of flex order pickers in the distribution centres. In Section 2.1.3 models from the most comparable application areas are discussed. Section 2.1.3 concludes the main findings on workforce management and uncertainty.

2.1.1. Workforce Management

Workforce management (WFM) involves matching workload with workforce (Nilssen, Stølevik, Johnsen, & Nordlander, 2011). The WFM process is divided into multiple stages. Some papers define four stages, whilst others identify only three stages. This is depicted in Figure 8, in which three papers are chosen as examples. Those three papers are not the only papers addressing this issue and making this distinction in stages. Within the following subsections, the four stages are briefly described.

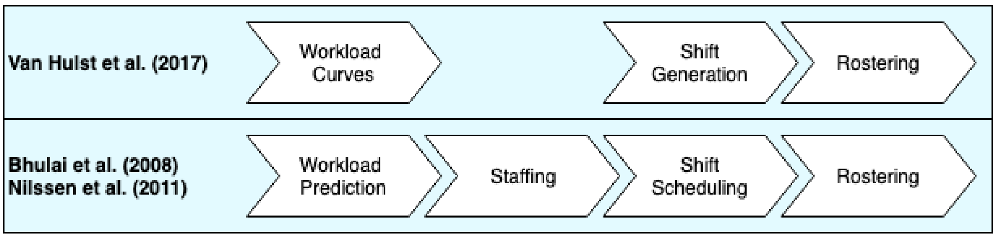

Figure 8: WFM Stages (van Hulst, et. al., 2017) (Bhulai, et. al., 2008) (Nilssen, et. al., 2011)

Workload Prediction

The term workload prediction indicates that this stage is used to determine the future amount of work (Bhulai, Koole, & Pot, 2008). However, Nilssen et al. (2011) already recalculate the amount of work into the demand for personnel. Finally, Van Hulst et al. (2017) describe the stage as the process of translating the workload information into workload curves to be used as input for the shift generation process.

An important aspect Van Hulst et al. (2017) note is the fact that there may be some uncertainty in the workload prediction. In an example of an Air Traffic Controller workforce planning problem they show that it is important to take the uncertainty of this prediction into account when generating shifts in the successive stage.

Staffing

The three papers all describe that staffing concerns the long-term management decision of how many staff to hire. Van Hulst et al. (2017) argue that the staffing stage is not a stand-alone process, since this is either a direct consequence of the shift design, or a strategic decision that must be made even before the first stage.

Shift Generation or Scheduling

The process of shift design consists of generating a set of shifts that fulfils requirements on minimum and maximum shift duration, legal start and end times, etc. The shifts are generated with an associated staffing demand assigned to each shift, so that the time-dependent and skill-dependent demands are satisfied at all time periods (Nilssen, Stølevik, Johnsen, & Nordlander, 2011). The objectives of the shift generation are to minimize the over- and underutilization, the total number of generated shifts, and the number of different shift types used (van Hulst, den Hertog, & Nuijten, 2017).

Most mathematical models used in practice for shift generation use estimated workload curves without taking the previously mentioned uncertainty in workload into account. This may lead to shift generation plans that are optimal for the estimated workload curves, but that are much less efficient for other workload realizations (van Hulst, den Hertog, & Nuijten, 2017).

Rostering

Based on the identified shifts as defined in the previous stage, the rostering stage assigns the personnel to the shifts. The rosters must fulfil all sorts of restrictions such as contract hours, employee preferences, and labour law regulations (van Hulst, den Hertog, & Nuijten, 2017).

Musliu et al. (2004) however, describe that there exist two main approaches in the literature to solve the shift generation and rostering. One of the approaches is to coordinate the design of the shifts and the assignment of the shifts to the employees, and to solve it as a single problem. The other considers the scheduling of the actual employees only after the shifts are designed, as shown in Figure 9 (Musliu, Schaerf, & Slany, 2004).

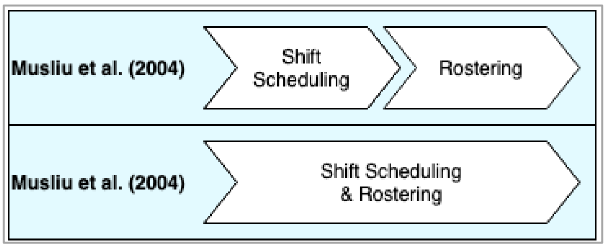

Figure 9: Two Types of Shift Scheduling and Rostering (Musliu, Schaerf, & Slany, 2004)

2.1.2. Uncertainty in Workforce Management

An important element of the workforce management of the operational personnel in the DCs, is the fact that the workload is uncertain. As already addressed in the previous section, uncertainty in the workload prediction influences the successive stages in the workforce management process. Mul et. al. (2006) categorize uncertainty into two groups: environment uncertainty and system uncertainty. The first, environmental uncertainty, includes uncertainties beyond the production process, such as demand and supply uncertainty. On the contrary, system uncertainty is related to uncertainties within the production process, such as operation yield uncertainty, quality uncertainty, and failure of production systems (Mul, Poler, García-Sabater, & Lario, 2006)

Both types of uncertainty occur in the DCs of Albert Heijn related to the flex order picking employees. The uncertainty in the productivity of order pickers is a type of system uncertainty. Whilst environment uncertainty is present due to the uncertainty in workload and unexpected illness of Albert Heijn employees. The uncertainty in workload is common in multiple application areas of workforce management and has been addressed in literature extensively.

The paper by Ernst et al. (2004) gives a comprehensive overview of the application areas in which research was performed concerning workforce management. They identify ten industries, or application areas, namely: transportation systems, call centres, health care systems, protection and emergency services, civic services and utilities, venue management (e.g. ground operations at an airport, or managing casinos and sport venues), financial services, hospitality and tourism, retail, and manufacturing. Due to the unique characteristics of those different industries and organisations, different types of models are required (Ernst, Jiang, Krishnamoorthy, & Sier, 2004). The most common application areas from literature, are discussed in the following subsections. The application areas and proposed solution models are described and their applicability to the problem at hand is assessed.

Transport Systems

Within the transport systems application area, airlines, railways, mass transit, and buses are comprised. Most models share two common features. First, that both temporal and spatial features are important, and second, all tasks to be performed by employees are determined from a given timetable (Ernst, Jiang, Krishnamoorthy, & Sier, 2004). Since the workforce is assigned to the predefined timetables, such as bus schedules or rostered flights, those models do not include uncertainty in demand. For example, busses will drive their regular roundtrip, even if no passengers are present at certain points in the trip.

The important spatial features of rostering personnel to bus lines or flights, are not present in the problem at hand of workforce management within DCs. However, the workforce within the DCs is dependent on the timetables of arriving and departing trucks. Yet, the importance of the spatial features within those transportation models makes those models unrelatable to the problem at hand.

Aircraft Services

In contrast to the spatial importance of models in transport systems, the venue management sector concerns the allocation of tasks at the same location. The operations often involve the completion of tasks with a variety of skill requirements. Examples of such problems include the ground operations at airports, cargo terminals, casinos, and sporting venues. The largest number of published papers in this application area are airport related staff scheduling problems. All of these problems are characterised by the fact that the demand for services is relatively well known as it is driven by the regular airline timetables, comparable to the transport system problems, however not restricted by spatial features (Ernst, Jiang, Krishnamoorthy, & Sier, 2004). This makes the models more applicable to the problem at hand.

However, the aircraft services are dependent on the actual arrival times of the aircrafts that are expected in a certain time period. This makes that there is almost no uncertainty in the total workload of a specific day. However, the uncertainty is in the arrival times of the aircrafts and thus results in uncertainty of workload at a given moment in time or the spread of the workload throughout the day.

Van Hulst et al. (2017) use Robust Optimization techniques to develop a shift plan for the Air Navigation Services, that is robust against this type of uncertainty in the workload prediction. This optimization technique means that the final shift plan may not be optimal for the estimated workload, but it is a very good plan for all possible realizations of the workload (van Hulst, den Hertog, & Nuijten, 2017). Whilst Hur et al. (2019) propose a model to deal with the uncertainty by using a rolling horizon break assignment procedure for the ground handlers at a major European airport (Hur, Bard, & Frey, 2019).

Since the shift at the Albert Heijn DCs are fixed, those types of optimization models are not applicable to the flex workers at the DCs. However, during the production days, the cockpit employees do use flexible break assignments. The cockpit employees base those decisions on the work that is already finished and that must be performed during the rest of the shift. Since the orders are already fixed, there is no need for a rolling horizon break assignment, such as required for the unexpected airplane arrivals.

Call Centres and Retail

Workforce management of call centres is an area for which an extensive amount of literature can be found. The literature review of Seada and Eltawil (2015) gives a good overview of those papers (Seada & Eltawil, 2015). Since call centres are not influenced by spatial difficulties like transport problems, call centres are also more comparable to the distribution centre problem at hand. However, call centres deal with highly random and uncertain demand (Liao, van Delft, & Vial, 2013).

The uncertainty in call centre models is the arrival of calls. Most call centre models in the literature assume that the calls arrive according to a Poisson process with known and constant mean arrival rates. However, data from practice often reveal that the process parameters are themselves subject to fluctuations (Liao, van Delft, & Vial, 2013). This level of randomness and uncertainty is not reached within the distribution centres, since the orders are fixed at the beginning of the production day.

Another addition common in the literature concerning call centres is the fact that call centres often have jobs that require different skills. This implies that the model should be able to deal with this multi-skilled complexity (Bhulai, Koole, & Pot, 2008). This is comparable to the situation in distribution centres, in which all employees can perform basic order picking tasks, whilst other tasks might require specific certificates such as forklift driving, order picking of medicines, or checking inbound goods. However, the research within this thesis is limited to flex order pickers only, thus multi-skilled models are not applicable.

Health Care

Multi-skilled models also occur in the nurse scheduling problem, as part of the health care application area. The majority of the literature on workforce management in health care concerns this multi-skilled nurse rostering (Ernst, Jiang, Krishnamoorthy, & Sier, 2004). The focus of the models in health care systems is the generation of rosters per nurse. This generation of shifts is hard due to tight regulations concerning night- and weekend-shifts. Staff preferences make the problems even more complicated, for example when preferred days off should be taken into account.

An advantage in nurse scheduling is the possibility at exploitation of economies of scale. In case two or more care units cooperate by jointly appointing a flexible nurse pool, the variability of random demand fluctuations balances out due to the economies of scale. If this principle is used, less buffer capacity is required to buffer against uncertainty (Kortbeek, Braaksma, Burger, Bakker, & Boucherie, 2015). Some DCs of Albert Heijn operate multiple departments, for example a non-perishable and perishable zone. It might be possible to apply similar economies of scale models as used in nurse scheduling by using those different zones.

Civic Services such as Postal Services

Civic services are services offered by the government. This can be at all levels, local, state, or national. Examples included claims processing, toll collection, and postal services. Of those, postal services are widely studied. The task in most cases is to create weekly schedules, with daily updates if significant changes in input parameters are identified. Many factors affect the behaviour of clients coming to the offices, for instance, weather conditions, holiday, days of the week, are some of the most important factors affecting the clients (Simeunović, Kamenko, Bugarski, Jovanović, & Lalić, 2017). Again, this type of uncertainty is not comparable to the uncertainty at the distribution centres.

Other Application Areas

Protection and emergency services are distinct to the problem at hand. The sector must deal with high service standards and tightly controlled regulations specifying acceptable patterns of shift work (Ernst, Jiang, Krishnamoorthy, & Sier, 2004).

2.1.3. Models in Comparable Application Areas

Table 2 summarizes the most important features of the application areas and models as described in the previous subsection. The table describes features of the workforce management process, and the applicability to each application area. In case the applicability of another application area matches the operational workforce management in DCs, the feature is highlighted in green.

Table 2: Feature Comparison of Workforce Management Application Areas

Feature	Operational Workforce in DCs	Transport Systems	Aircraft Services	Call Centres	Health Care
Fixed Timetable Driven	No	Yes	Yes	No	No
Spatial Features	No	Yes	No	No	No
Multi-Skilled	Yes	No	No	Yes	Yes
Uncertainty	Daily Demand	N.A.	Spread throughout day	Daily Demand	Daily Demand
Uncertain Input Becomes Fixed	Before Execution (#colli) & During Execution (productivity)	During Execution	During Execution	During Execution	During Execution
Uncertainty Level	High	N.A.	Low	High: Poisson	High But, might exploit economies of scale

The call-centre and health care models seem most comparable to workforce management at distribution centres. However, there are still some major differences concerning the level of uncertainty and the time the realized values become known.

Examples of successful models to deal with uncertainty in call centres include the research of Liao et. al. (2013). They use estimated seasonal and global busyness factors of call centres from past data, to predict the uncertain arrival rates. By combining stochastic programming and distributional robust optimization, they aim to minimize the total salary costs under service level constraints. The combination of those models make it possible to build a trade-off curve between the salary costs and various measures of satisfaction, e.g., average number of times the constraint is satisfied, conditional expectation of the understaffing, or maximum understaffing (Liao, van Delft, & Vial, 2013).

In the research of Pakpoom & Charnesethilkul (2018), uncertainty in workload predictions is taken into account, by using the two-stage stochastic integer program. Those programs can be solved by the CPLEX MIP solver. However, Pakpoom & Charnesethilkul (2018) applied Benders decomposition and derived a special solution method to solve the two-stage scheduling problem. They generated 16 test instances, all for which the algorithm considerably outperformed solving by the CPLEX MIP solver. The results showed that the proposed algorithm can reduce time to solve to optimality by more than tenfold. However, the proposed method only works on cases whose possible demands are even, and total number of time periods is odd (Pakpoom & Charnesethikul, 2018).

2.1.4. Conclusion

Although uncertainty in workforce management is a commonly studied problem in multiple sectors, most literature is not relevant to our problem. Most industries deal with different types of uncertainty, e.g., arrival times of customers, the number of customers, or the time required per customer. Those uncertain variables only become known during the execution. However, the workload at the DCs of Albert Heijn become known shortly before the start of the first shift.

Furthermore, most proposed solutions aggregate multiple phases of the workforce management process, for example combining workload prediction and shift generation. However, the method used at Albert Heijn with fixed shifts and a lot of responsibilities placed at the employment agencies, results in the fact that only the workforce prediction phase of the workforce management process is relevant.

2.2. Data Mining for Predictions

Within this section relevant literature concerning data mining is addressed. In the first section, Section 2.2.1, previous research of data mining in the field of workforce management is addressed, with the focus on the most comparable application areas as identified in the previous section. In Section 2.2.2 the relevance of prediction intervals and appropriate methods to determine those intervals are discussed. Section 2.2.3 elaborates on different accuracy measures used in forecasting. The main findings from this section are summarized in Section 2.2.4.

2.2.1. Data Mining Applied in Workforce Management

A lot of the research in predicting workload concerns the usage of Neural Networks (NN), or Artificial Neural Networks (ANN). For example, Millán-Ruiz et. al. (2010) focus on forecasting call arrivals. They analyse and compare ANN, two Time Series models: Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing (ES), and a Lineal Regression Model. They compare the different models for five types of incoming calls. The best forecast method for all call types is the ANN. This model results in the smallest MAE and standard deviation, for example the MAE is almost 13% lower in case of the ANN compared to the regression model for call type 1. The reasoning they provide is that, while ARIMA and Time Series techniques focus on the recent past, ANNs are more flexible for longer time horizons (Millán-Ruiz, Pacheco, Hidalgo, & Vélez, 2010).

Neural Networks also performed well in the research of Serengil & Ozpinar (2017). They compared NN and exponential smoothing algorithms for forecasting workload for bank operation centres. The results of the exponential smoothing methods cannot get close to the NN results. The NNs have a Mean Absolute Error (MAE) of 71, which is 14% of the Mean, whilst the exponential smoothing achieves an MAE of 284 which is 56% of the mean (Serengil & Ozpinar, 2017).

The research of Ruiz-Aguilar et. al. (2017) also successfully used ANNs. The research focused on prediction of the daily number of goods, subject to inspection at Border Inspections posts (BIPs). Different parameter settings used in ANNs with Bayesian regularization outperformed the Multiple Linear Regression models (Ruiz-Aguilar, Moscoso-López, Turias, & González-Enrique, 2017).

In addition to ANNs, decision tree like models are used as well. For example, Gomes et. al. (2012) compare Linear Regression (LR), Random Forest (RF) and the M5 Rules (M5) algorithm, which is a type of a decision tree learning model, to reduce the uncertainty of the estimated duration of surgeries. The results show there is a clear advantage of using the data mining algorithms for predicting surgery duration, when compared to the estimates made by surgeons. The estimation accuracy is improved by almost 36% by the M5 algorithm. Thus, one could assume that the M5 algorithm provides 36% more time to perform surgeries. The LR and RF models also significantly improve the accuracy (Gomes, Almada-Lobo, Borgers, & Soares, 2012).

The research by Gomes et. al. (2012) showed Linear Regression models performed well. However, those models only allow the description of a continuous, symmetric response in terms of a linear combination of prediction variables. Generalized Linear Models extend this framework to a wider range of response types, including categorical, binary, and skewed continuous responses (Faraway, 2010).

In the research of Gianazza (2017), multiple models are assessed to predict the workload of air traffic controllers. This prediction is one of three options: low workload, normal workload, or high workload. To predict this class, the Neural Network performed best, followed by Gradient Boosted Trees. The models realized an 81,9% and 81,8% accuracy when predicting the classes. The other models analysed in the research only reach 77% or lower accuracies. Those models are Naïve Bayes Classifiers, Linear Discriminant Analytics and Quadratic Discriminant Analysis (Gianazza, 2017).

Since the literature above clearly shows Neural Networks are widely and successfully used in the prediction of workload, we will also apply this method in this research. Additionally, two tree-like models will be used: Random Forest and Gradient Boosted Trees. Finally, both Generalized Linear Models and normal Linear Regression models will be used within this thesis.

2.2.2. Prediction Intervals

For many real-world applications, it is not enough that on average a model performs well, rather the uncertainty of each prediction must also be quantified. Prediction Intervals (PIs) directly communicate uncertainty, offering a lower and upper bound for a prediction and assurance that, with some high probability (e.g. 95% or 99%), the realised data points will fall between these bounds. Having this information allows for better-informed decisions (Pearce, Zaki, Brintrup, & Neely, 2018).

To produce a prediction interval, it is necessary to have an estimate of the standard deviation of the forecast distribution, denoted by $\hat{\sigma}$. When forecasting one step ahead, this standard deviation of the forecast distribution is almost the same as the standard deviation of the residuals. Equation 1 shows how this residual standard deviation can be calculated (Hyndman, 2018).

However, for a multi-step forecast, a more complicated method of calculation is required which assumes that the residuals are uncorrelated. One of those methods is called the mean forecast, in which the standard deviation of the h-step ahead forecast, denoted by $\hat{\sigma}_h$, is calculated by Equation 2. Equation 3 shows how the prediction interval for the h-step forecast can then be calculated.

$$Residual\ Standard\ Deviation = \hat{\sigma} = \sqrt{\frac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{T-2}} \quad (1)$$

$$\hat{\sigma}_h = \hat{\sigma}\sqrt{1 + \frac{1}{T}} \tag{2}$$

$$\hat{y}_{T+h|T} \pm t_{\frac{\alpha}{2},n-1} * \hat{\sigma}_h \tag{3}$$

This method of defining the prediction interval is applicable to all types of prediction models. However, Montgomery and Runger (2011) define a slightly adjusted method to define prediction intervals for regression models. This method includes the value of the regressor variable used to make the prediction (Montgomery & Runger, 2011).

In case of a single linear regression model, a feature Y can be predicted by a regressor variable x. In case of the variable x_0, the point estimation for the response Y_0 is then given by Equation 4. The $100(1-\alpha)\%$ prediction interval on the future observation Y_0 at the value x_0 is given by Equation 5, in which $\hat{\sigma}^2$ denotes the mean squared error of the residual error, n the number of observations used to determine the regression model, $\bar{x}$ the average value of the regressor variable x, and $S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$ the sum of the squares of the difference between each x and the mean x value (Montgomery & Runger, 2011).

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 \tag{4}$$

$$\hat{y}_0 - t_{\frac{\alpha}{2},T-1}\sqrt{\hat{\sigma}^2\left(1 + \frac{1}{T} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)} \leq Y_0 \leq \hat{y}_0 + t_{\frac{\alpha}{2},T-1}\sqrt{\hat{\sigma}^2\left(1 + \frac{1}{T} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)} \tag{5}$$

For linear regression models with multiple regressor variables $(x_1, x_2, \ldots)$, the determination is given by Equation 6 (Dean, 2016).

$$PI = \hat{y} \pm t_{\frac{\alpha}{2},T-1}\sqrt{MSE + CovInt_i + 2\left(\sum x_i * CovInt_i\right) + \left(\left(\sum x_i * CovX_i\right) * x_i\right)} \tag{6}$$

2.2.3. Assessing Forecast Accuracy

When comparing different forecasting methods, different measurements can be used. Within this section three types of accuracy measures and their advantages are described.

The first method to measure the accuracy of prediction methods is the Mean Absolute Error (MAE), given by Equation 7. The absolute error is used, since averaging both negative and positive errors can result in an average error close to zero in case the negative and positive errors are similarly different from zero.

$$MAE = \frac{1}{T_f} \sum\nolimits_{t=1}^{T} |y_t - \hat{y}_t| \quad (7)$$

The MAE is a scale-dependent accuracy measure, since it is in the same scale as the dependent variable. Thus, in case the performance of models concerning different dependent variables should be assessed, the MAE is of no value. However, the Mean Absolute Percentage Error (MAPE) rescales the values of the MAE in the interval [0,1] (Gomes, Almada-Lobo, Borgers, & Soares, 2012). The MAPE is calculated by Equation 8.

$$MAPE = \frac{1}{T_f} \sum_{t=1}^{T} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (8)$$

A third method to determine the accuracy is the Root Mean Squared Error (RMSE), calculated by using Equation 9. In which y_t and $\hat{y}_t$ represent the actual and forecasted values at time t respectively. And T is the number of observations in the dataset which used to test the accuracy. By using the quadratic loss function, the RMSE weight under and over-estimation of the same magnitude in the same way, of which the importance is already shown in the example above (Bou-Hamad & Jamali, 2020).

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T} (y_t - \hat{y}_t)^2}{T}} \quad (9)$$

The RMSE measure is more appropriate in case small errors are less relevant. Due to the squared error, large errors have a disproportionately large effect on the RMSE outcome. Furthermore, the RMSE is similar to the MAE in the sense that it is a scale-dependent accuracy measure, making it not applicable in comparing the performance of different dependent variables.

2.2.4. Conclusion

Within this section the application of different data mining techniques in workload prediction are discussed. Based on the state of the art in literature, five prediction methods are selected to be used within this research: Neural Networks, Random Forests, Gradient Boosted Trees, Generalized Linear Models, and Regression Models. Instead of using point forecasts, methods are described to generate prediction intervals for the prediction models. To assess the performance of the different model types, the Mean Absolute Percentage Error (MAPE) is the most applicable method to compare the accuracy.

3. Context Analysis

Within this chapter, the current practices of Albert Heijn are described concerning the workforce management of operational personnel, specifically the order pickers, at the DCs. To do so, multiple interviews were held, which is explained in the first section of this chapter. The following sections describe the different phases of the workforce management process. While describing the processes, hypotheses are drawn based on the interviews, previous research at Albert Heijn, and the literature review of Section 2. Those hypotheses are tested in Chapter 4, by analysing the available data.

The hypotheses are either relevant to the next step, the data mining, since they might identify insights in the data which could be exploited. Otherwise, or additionally, the hypotheses are relevant for Albert Heijn, to identify existing questions concerning the data. Are some of the feelings of the employees indeed true? Or are some assumed improvement points not relevant to improve?

3.1. Methodology

To get an insight in the current way of working, multiple interviews were conducted in which the interviewee was asked to show his default way of working concerning the workforce management. Appendix A lists all those conducted interviews. Since Albert Heijn operates multiple DCs, it was important to get an insight in the methods of all distribution centres. However, multiple actors are part of the WFM processes. Due to time constraints it was not feasible to interview all those actors at each DC. Therefore, only the most important actor, the capacity planner, is interviewed at each DC.

During those meetings with the capacity planners, they were asked how they perform their tasks and to show this in real life. Those were unstructured interviews, since each capacity planner has their own way of working and preferred sequence in how to explain this process. Some of the capacity planners where very brief, whilst others could fill at least two hours with explaining and examples. In addition to describing the current processes, each interviewee was also asked beforehand to think of possible improvement opportunities, which were discussed during the interview as well.

In addition to capacity planners, shift leaders and cockpit employees are interviewed using the same method. Due to time constraints only one DC was selected to do so. Since the employees of the DC in Zwolle were very cooperative, this DC was selected to interview those employees. Furthermore, a brainstorm was held to describe the current way of working and identify improvement opportunities. The details concerning this brainstorm are given in Appendix B.

Additional meetings where arranged with less directly related employees such as IT developers and employees of the Replenishment and Logistics Preparation department, to get an insight in related processes such as forecasting and data storage. Finally, two meetings organized by the Supervisor Flex employees where attended. Those supervisors organize those meetings every two weeks to discuss issues concerning the employment agencies of each DC. During the attend meetings, the focus of the meetings lied on the update in the WAB and how to deal with this.

3.2. Workforce Management at DCs of Albert Heijn

In Section 2.1, the four stages of workforce management as defined in literature were described: workload prediction, staffing, shift scheduling, and rostering. Although slightly adjusted, those stages can be recognized in the workforce management process at the DCs of Albert Heijn, which is shown in Figure 10. The figure schematically shows which actors are responsible in a certain state and what they are responsible of to deliver. The different phases are described below.

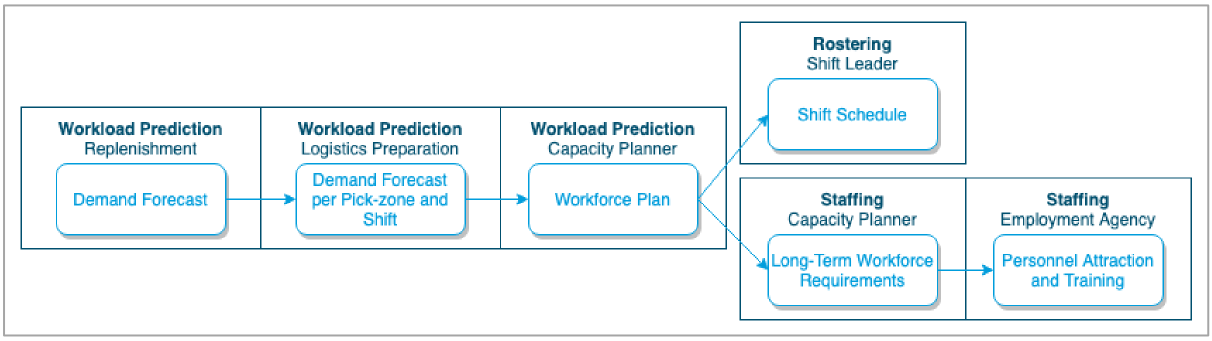

Figure 10: Workforce Management of Operational Personnel at Distribution Centres

Workload Prediction

The workload prediction is done by the Replenishment department, by predicting the total number of colli to pick during an entire production day. A detailed description of their prediction method is given in Section 3.3.1. This total per production day is translated into a forecast per pick zone and shift by the Logistics Preparation department, based on historical data. This information per pick zone and shift is in turn used by the capacity planner, who alters the prediction based on the expected service percentage of the DC, to finally determine the required number of employees per shift. Those calculations performed by the capacity planner are described in Section 3.3.2.

Staffing

Each DC has a pool of employees employed by Albert Heijn, from now on abbreviated to AH employees. Those employees are either full-time or part-time employees. The part-time employees mainly work during the weekend shifts, whilst the full-time employees mainly work during the week. Around 2012, the management has decided that AH is not hiring full-time employees anymore. Only some part-time employees are hired if necessary, to fulfil weekend and night shifts. However, since this concerns such a small number of employees, the staffing phase is assumed to be not relevant for AH order pick employees.

The workload cannot be covered by AH employees only. Therefore, the hiring of flex employees is crucial at each DC. Although the employment agencies are responsible for their own staffing processes, input given by the capacity planner is of extreme importance. For example, during the summer or Christmas period, the turnovers are higher, requiring additional flex employees. In case the employment agencies must fulfil this demand, they must be informed in time to attract and train new personnel to cover the peeks.

Shift Scheduling

Each DC has fixed shifts. The shifts are fixed since this is most convenient to work with. Employees are used to their start and end times, and the shifts are designed in such a way it matches the in- and outbound truck schedules. Thus, shift scheduling is not an operational problem in the workforce management of Albert Heijn.

However, circumstances at a DC might change, sometimes requiring changes in shifts, making the shift schedule a tactical decision. For example, the DC in Zaandam is being automized by slowly transferring the delivery of goods from the original manual warehouse, to the automized warehouse. The workload of the manual warehouse therefore gradually decreases over time and the number of inbound and outbound trucks decrease as well. Thus, to prevent employees without tasks, the shifts are adjusted throughout the automation process. Another example of changing the shifts is at DC Zwolle. Currently they work with a night and day shift to pick all colli, and during the evening shift the DC is prepared for the new production day. Due to the growth of the total volume required from the DC, the management would like to test what happens if the evening shift will also perform order picking tasks, to reduce the number of orders to pick during the night shift.

Rostering

The rostering is performed by the team leaders. Once the capacity planner has determined the required workforce, those hours are filled in Interflex. Interflex[1] is a WFM system used by Albert Heijn to manage the hours of AH personnel. The system is able to show a Gantt chart, displaying a block for each task with a start and end time. The team leaders can assign employees to the specified task blocks. Once a block is assigned to an employee, the colour changes. The capacity planner can check whether all tasks are fulfilled. In case not all hours can be filled, the capacity planner is responsible to fill those missing hours with flex workers.

[1] https://www.interflex.nl/nl/index.html

3.3. Workload Prediction and Staffing

This section describes the workload prediction and staffing phases of DCs. The order picking workload is dependent on the total number of colli to pick. Therefore, the workload prediction starts with the colli prediction. This colli prediction process is described in the Subsection 3.3.1. Subsection 3.3.2 subsequently describes how this colli prediction is translated into the prediction of the required number of order pickers.

3.3.1. Colli Predictions

The colli prediction is the first step in the workforce prediction process. This section describes the different colli prediction moments and how those predictions are made.

The colli predictions are made by the Replenishment department. The department releases multiple updates of the prediction. This is shown in Figure 11. In the figure, the production day is denoted with X. The first prediction is given seven weeks in advance, denoted with X-7. An update is made two weeks and one week in advance, denoted by X-2 and X-1 consecutively. The different prediction updates and their importance are described below.

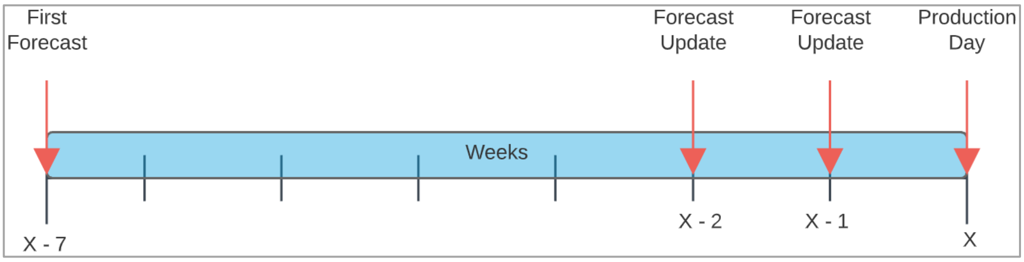

Figure 11: Prediction Update Moments

Prediction at X-7

The first prediction is based on experience and data of previous years, combined with the promotional actions that are expected. It is not possible to anticipate all promotional actions seven weeks in advance. For example, a last-minute change can be made by the marketing department in case the weather predictions change to perfect BBQ weather. Other causes of varying promotional actions less than seven weeks in advance could be the promotions of competitors, or the availability of resources.

Based on the predicted number of colli, the capacity planner can determine the required workforce as will be explained in the following section. Based on this required workforce, the capacity planner can inform the employment agency of the expected required number of flex workers. This first indication of required hours is given to the employment agency seven weeks in advance since the employment agency is then still able to fulfil their staffing task. They have time to attract and train new personnel if required or downscale the number of flex workers in the pool by allowing holidays of the flex workers.

Customer, Store, and Warehouse Demand Forecast

From four weeks in advance, the Replenishment department starts forecasting based on a customer centred view. Figure 12 schematically depicts how the customer is the starting point, and the predicted demand influences the demand higher up in the supply chain. In general, the Replenishment department uses the "out = in" approach, meaning that each sold item should be restocked as soon as possible. Thus, the customer demand forecast (CDF) of a product at a certain store, determines the order quantity of that certain product at that specific store.

Table 3: Example Calculation CDF

Forecasted Variable	Notation / Equation	Example Value
	confidential	

In general, the forecasted number of customers is simply the number of customers for the same day in the previous week, adjusted for weather or special events, e.g., Easter, Christmas, etcetera. Manual adjustments are made if required, either in the in- or exclusion of historical days, or the forecasted number for N. The prediction of the demand per item per customer (PQ) is somewhat similar. A simple algorithm is used to predict the values based on the sales transactions over the past 3-12 weeks. This algorithm also takes the promotional actions into account.

Once the demand per product is predicted based on the prediction of the number of customers and the items per customer, the store demand forecast (SDF) is simply the sum of the forecast of all products of that certain store. The SDF is thus the overall number of colli that is expected to be ordered by the store. This forecast is most important for the distribution centres, since the sum of all store demand forecasts of stores delivered from a specific DC, determines the total number of colli to pick in the DC.

To complete the supply chain process, the sum of store demand forecasts per distribution centre determines the warehouse demand forecast (WDF), which is used to determine the required replenishment orders from suppliers to the distribution centres.

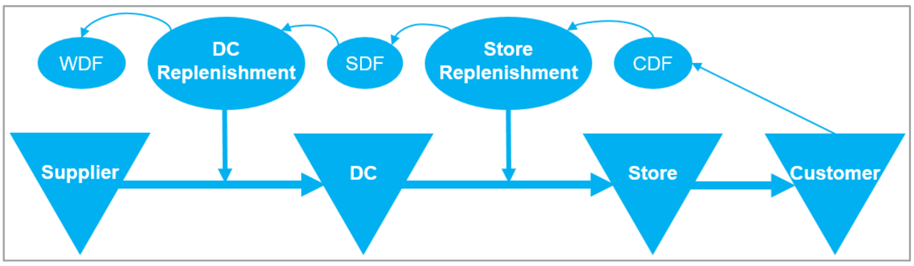

Figure 12: Prediction Process of the Replenishment Department

Prediction at X-2

Although the CDF, and thus SDF, are made four weeks in advance, the first prediction update given by the Replenishment department is two weeks in advance, the X-2. This prediction is mainly used by the Transport department to determine their truck schedule. Based on the demand per store, the best combinations of stores in trucks can be made to minimize the total number of trucks required. However, each store has a timeframe in which the delivery should take place, making this problem harder to solve.

Off course, the capacity planner will also use the updated prediction to recalculate the required workforce. In case there are very large differences compared to the prediction given at X-7, the employment agency is informed. Furthermore, the team leaders will now start planning their employees in Interflex, as described in the rostering step.

Prediction at X-1

Finally, one week in advance the last prediction is given. This prediction is most important for the DCs, since this is the last prediction that can be used to adjust the requested number of flex workers at the employment agency. It is known that the prediction given at X-1 is the most accurate as well, since this prediction is closest to the actual production day. This is also shown in Table 4, which shows the mean absolute deviation, MAD, per pick zone and shift for X-1, X-2, and X-7. The lower the MAD, the better the forecasted value. As the table shows, the MAD is lowest at X-1 for all pick zone and shifts. Thus, the forecast error declines when the forecast horizon gets smaller.

Table 4: MAD Forecast X-1, X-2, X-7, per Shift and Pick Zone, DCO, 2019

	Non-Perishable		Perishable A	
	Day	Night	Day	Night
MAD X-1				
MAD X-2	*Confidential*			
MAD X-7				

Prediction Moments

Since all predictions (X-7, X-2, X-1), are only given at Wednesdays, this implies that for the prediction of Sunday, the last update is more than one and a half weeks in advance of the actual production day. This is depicted in Figure 13. Multiple of the interviewees mentioned this as a possible improvement point. It was also one of the improvement points identified during the brainstorm. It is expected that with updating the prediction more often during the week in advance of the production, the forecast could be more accurate. This leads to the first hypothesis:

Hypothesis 1: The absolute deviation between the predicted and realized number of colli increases throughout the week due to the non-rolling prediction after the last update.

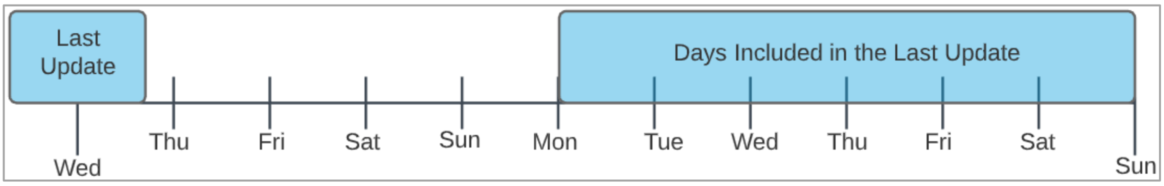

Figure 13: Last Colli Forecast Update

The first hypothesis could help in deciding whether to invest in creating a rolling forecast. However, this feeling of an increasing error throughout the week also suggests a weekly cyclic pattern of the forecast error. If such a weekly pattern exists, this could be exploited in the update prediction method.

In addition to a weekly pattern, some employees also expect some broader patterns, for example three or four weekly or monthly. The Replenishment department adjusts their predictions based on the previously realized demand. In case overestimations where made in the previous days, the Replenishment department can downscale the predictions of the upcoming days. However, if this downscaling is done to heavily, the following days will be underestimated. In case such a cyclic pattern can be found in the predictions, this could be exploited in different model types. Thus, the second hypothesis aims to identify any cyclic pattern in the prediction error.

Hypothesis 2: The prediction error of the X-1 prediction follows a cyclic pattern.

Finally, in case there exists a correlation between the forecasted volume and the error of the forecast, this knowledge could also be included in the models to determine the required number of flex employees. Resulting in the third hypothesis.

Hypothesis 3: The forecast error is correlated to the forecast volume.

Box-Control

Within the DCs, the prediction given at X-1 is the key input for the personnel request to the employment agency. However, the exact demand from stores can still vary significantly. To reduce the difference between the prediction at X-1 and the realized demand from stores, the Replenishment department works with box-controlling. By classifying products into different priority groups, the Replenishment department can easily adjust the actual amount of colli that must be delivered from a DC.

The main classes in box-controlling are “must”- and “may”-colli. A product is classified as must-colli, in case the store will run out of stock if the product is not replenished during the upcoming production day. A product classified as may-colli, is a product which is most likely to run out of stock at the supermarket in the near future, however, this will not happen before the second delivery moment.

In case the orders of stores resulted to be higher than predicted, the Replenishment department can decide to cancel some of the may-colli in the delivery. The total amount of colli that must be delivered from a DC will thus be reduced. On the other hand, in case less products are ordered than expected, the Replenishment department might decide to ship additional may-colli, to increase the total number of colli.

However, deciding not to deliver the products classified as may-colli, will most likely result in must-colli in the subsequent days. For example, a specific type of peanut butter was classified as may-colli for the production day Tuesday, since there will still be sufficient stock to cover demand before the delivery at Wednesday if no replenishment takes place. However, on Wednesday, the store will run out of Peanut butter in case no replenishment takes place, as shown in the example in Figure 14. Therefore, postponing the delivery of may-colli, results in must-colli. Although this example shows the effect on the direct subsequent day, this might also take another day.

Replenishment Day	**Monday**	**Tuesday**	**Wednesday**
Stock at Store at Decision Moment	15 Peanut Butter	10 Peanut Butter	5 Peanut Butter
Expected Demand before Upcoming Replenishment	5 Peanut Butter	5 Peanut Butter	5 Peanut Butter
Resulting Stock when Replenishment will Arrive	10 Peanut Butter	5 Peanut Butter	0 Peanut Butter
Decision	**No Order**	**May-Colli**	**Must-Colli**

Figure 14: Explanation Basic Box-Control

Thus, although box controlling is a good method for the Replenishment department to control the actual number of colli to deliver, this method cannot be used to exactly match predicted and realized orders. Therefore, realized orders still deviate from the predicted colli values.

Service Percentage

The last important note on the colli predictions made by the Replenishment department, is the fact that the realized orders, are not directly the realized number of colli picked. For example, it is possible that products are not in stock in the distribution centre at the moment of order picking. Thus, the realized orders are also known as the "gross colli". The actual number of picked colli, is know as the "nett colli".

The difference between the gross and nett colli is called the "service percentage" of the distribution centre. A service percentage of 100% indicates that all ordered products, are available and thus picked. However, the service percentage at DCO in the non-perishable pick zone is generally around 96-98%, indicating that approximately 2-4% of the ordered colli are not picked due to unavailability in the distribution centre.

The calculation to determine the service percentage is shown in Equation 10. The same calculation is shown in Equation 11 with the different terminology of gross and net colli.

$$Service\ Percentage = \left(1 + \frac{Picked\ Colli -\ Ordered\ Colli}{Ordered\ Colli}\right) * 100\% \tag{10}$$

$$Service\ Percentage = \left(1 + \frac{Net\ Colli -\ Gross\ Colli}{Gross\ Colli}\right) * 100\% \tag{11}$$

Exploiting Economies of Scale

The literature review described that some workforce management models exploit economies of scale in determining the required workforce. An example was the usage of flexible assignable nurses used in the planning of two distinct nurse wards. In case two or more care units cooperate by jointly appointing a flexible nurse pool, the variability of random demand fluctuations balances out due to the economies of scale. Since most distribution centres of Albert Heijn consist of multiple pick zones, the models might be applicable as well. For example, an increase in the number of colli to pick in the non-perishable zone, might be related to a decrease in the number of colli in the perishable pick zone. If this is the case, a flexible worker pool could be used similar to the flexible nurse pool. To assess this, the fourth hypothesis is:

Hypothesis 4: There exists a negative linear correlation between the forecast deviation in the non-perishable and perishable A pick zone.

3.3.2. Translating Workload to Required Workforce

The capacity planner is responsible to translate the given workload information obtained by the Replenishment and Logistics Preparation department into a required number of employees, the size of the workforce. The capacity planner determines this number at the three moments that the Replenishment department updates the colli prediction for the production day. As described in the previous section, those moments are 7, 2, and 1 weeks in advance.

As explained in the introduction, each DC has a fixed pool of AH employees. Therefore, the main question in determining the required workforce is the required number of flex employees. This depends on the total workload and the available AH employees. This section describes the calculations required to determine the number of flex order pickers.

The Basic Calculation

The order picking task is called a "direct task". This is the case since the required hours of order picking are directly related to the required number of orders, or colli, to pick. The required hours of order picking can be calculated by using Equation 12. Note that this equation uses the nett number of colli to pick. Equation 13 shows that this is the same as the gross colli times the service percentage.

$$Required\ Order\ Picking\ Hours = \frac{\boldsymbol{Nett}\ Number\ of\ Colli\ to\ Pick}{Number\ of\ Colli\ Picked\ in\ One\ Hour} \tag{12}$$

$$Required\ Order\ Picking\ Hours = \frac{\mathbf{Gross}\ Colli * Service\ Percentage}{Number\ of\ Colli\ Picked\ in\ One\ Hour} \tag{13}$$

The number of colli picked in one hour is also called the productivity of the order pickers. From now on, we denote the productivity by PR. All capacity planners simply use a single value for the productivity. Some capacity planners base this value on the given budget made by the finance department, whilst others use historic data from the previous weeks or even years.

Thus, based on the productivity and number of colli to pick, the required hours of order picking can be calculated, this is also depicted in Figure 15. However, as the figure shows, to determine the required hours of flex workers, the illness rate and the days off of Albert Heijn employees should still be taken into account.

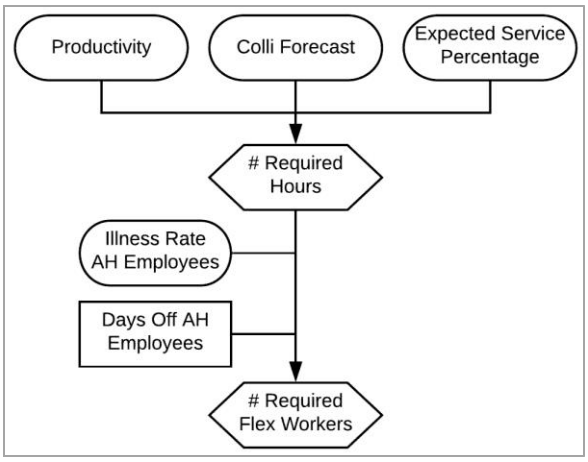

Figure 15: Workforce Prediction

Including Days Off

Some of the capacity planners have a detailed planning of the days each of the Albert Heijn employees will be absent for holidays, STWA (in Dutch ADV), or other free days. However, some of the capacity planners do not collect this information and are thus forced to estimate how many Albert Heijn employees will be available. For simplicity, within the rest of this thesis it is assumed that all capacity planners are aware of the free days of their employees, making this a known input parameter to the calculations. It is however, important to include in the recommendations that this is something which Albert Heijn could and should implement at all distribution centres to improve the planning process.

Absence due to Illness

Although it is now assumed that the days off from Albert Heijn employees are known, there is still variability in the actual number of own employees that will be working on the production day. This is caused by the possible absence of employees due to illness. The capacity planners use an illness rate based on historical data to determine the expected number of Albert Heijn employees that might be ill at the day of production.

Multiple capacity planners identified a difference in the illness rate per production day. Since most AH employees do not work during the weekends, the working week is from Mondays till Fridays. Therefore, most capacity planners feel like the Monday is the day at which the most unexpected illness is present. During the rest of the week, the team leaders speaks with their employees and might predict illness in case someone is not feeling well the day in advance.

However, although illness might be predicted more accurately during the weeks due to the direct contact with the team leaders, this is not four days in advance of the production day. Therefore, this information is not valuable in reducing the uncertainty in the workforce predictions. The total number of absent employees should therefore be estimated by the capacity planner. It is however expected that this number does follow some seasonal patterns. For example, more employees are ill in the fall and winter seasons, than during spring or summer. Or, at the beginning of the week, more employees are absent. Those assumption results in the following hypothesis.

Hypothesis 5: The absence of AH employees due to illness follows a seasonal and/or weekly pattern.

Based on this illness rate, the total hours of flex workers can be calculated by Equation 14. In this equation, the illness rate is denoted by I, which is a value between 0 and 1 indicating the fraction of hours of AH employees will be missing. C is the predicted number of colli, S% the expected service percentage, PR denotes the predicted productivity, and AH denotes the planned working hours of AH employees.

$$Required\ Flex\ Hours = \frac{C * S\%}{PR} - \left(AH * (1 - I)\right) \quad (14)$$

The calculation method shown in Equation 14 is the current method all capacity planners work with.

Productivity of Order Pickers

However, among all capacity planners it is known that the order pickers have different productivity rates. The flex workers have a higher productivity, since they will be financially rewarded if they achieve a certain threshold. Albert Heijn employees have a lower productivity, not only since they are not rewarded, but most likely since they are older and do not like the order picking tasks. Since this variability is known, it would be better to denote the productivity as PR_E, in which the subscript E denotes the employee type. The types that could be distinguished are flex workers and Albert Heijn employees. This results in the following hypothesis.

Hypothesis 6: The average productivity of Flex employees is higher than the productivity of AH employees.

In case the productivity per employee type should be taken into account, the simple calculation given in Equation 14 to determine the required hours of flex workers does not hold. The calculation should then be made in two steps. First, determining the number of colli that is expected to be picked by the flex employees, by calculating how much colli the AH employees can pick. This can be done as shown in the top part of the division in

Equation 15. The available AH hours, reduced by the expected illness, are multiplied by the productivity (PR_{AH}). This is the amount of colli the AH employees are expected to pick. This is subtracted from the total number of predicted colli to pick (C). Secondly, the productivity of the flex worker can then be used to determine the required hours, by dividing the resulting number of colli by the productivity of the flex employees (PR_{flex}).

$$Required\ Flex\ Hours = \frac{\mathrm{C} * \mathrm{S}\% - \big(AH(1-I)\big) * PR_{AH}}{PR_{Flex}} \tag{15}$$

Since the capacity planners are currently not including the differences between flex and AH employees, the following hypothesis is drawn:

Hypothesis 7: The prediction accuracy of the required workforce could be improved by taking the differences in order pick productivity into account between AH and Flex employees.

The current method uses the uncertain variables to determine the required number of flex employees. It is assumed that the colli predictions fluctuate the most, causing the biggest differences in required number of flex employees. If this is the case, the colli predictions are the most important uncertain variable to address when looking for improvements in the current prediction method. This results in the eighth hypotheses.

Hypothesis 8: The colli prediction has the largest relative error and is thus the most uncertain variable of the three uncertain variables.

Previous Research on Order Picker Productivity

Previous research on order picker productivity has been performed at Albert Heijn in 2019. The thesis by de Lang (2019) focusses on creating a framework of order picker productivity, shift scheduling, and staging lane usage. His most important finding concerning the order picking productivity is the fact that the productivity of order pickers decreases, in case the number of order pickers working at the same time increases.

Figure 16 shows one of the graphs as a result of de Lang (2019), in which the x-axis shows the number of order pickers active at the same time, and the y-axis the average productivity of the order pickers. A clear downward trend can be seen in the graph, indicating the reduction in productivity. De Lang concluded this decrease in productivity can have multiple causes such as longer break times due to waiting lines at coffee corners and more people to talk to, as well as reduced productivity due to congestion. In this case congestion is defined as waiting time of order pickers for example due to waiting for a pick replenishment, or simply waiting to pick from a certain location since another order picker is currently picking from the location (de Lang, 2019).

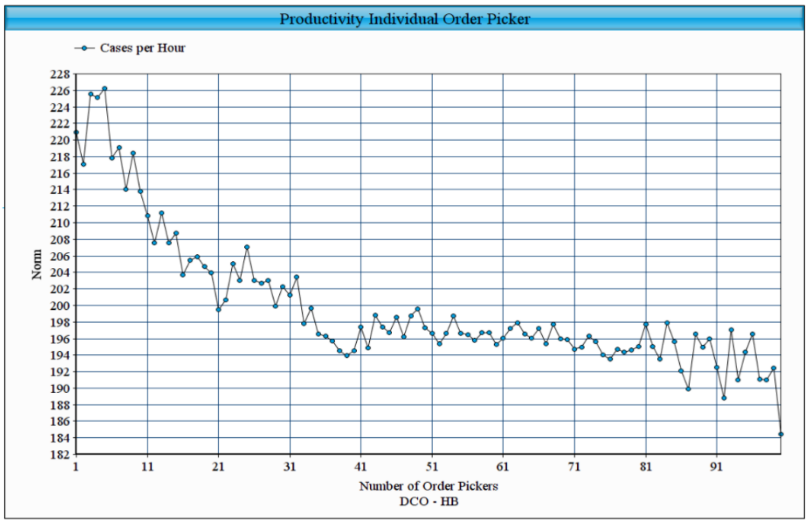

Figure 16: Order Picker Productivity vs. Number of Order Pickers (de Lang, 2019)

When predicting the expected productivity by data mining, the number of employees that will work is not known yet. Only the predicted amount of colli is known. Therefore, it is relevant to test if the same relationship holds between the average productivity and total number of colli to pick during a shift. This results in the last hypothesis.

Hypothesis 9: The average order picking productivity per shift is negatively linear correlated to the total number of colli to be picked in the entire pick zone per shift.

3.4. Conclusion

Based on the interviews, brainstorm session, previous research, and the literature review, the current way of working has been described within this chapter. Improvement possibilities are identified. Within this chapter those possibilities are translated to hypotheses, which will be assessed in the following chapter by analysing the available data.

In total, the context analysis resulted in eight hypotheses. Those hypotheses are shown in Table 5, together with a classification. Four of the hypotheses will mainly add value by giving insights relevant for the data mining task. The other three hypotheses are mainly valuable due to their possible improvement to the current way of working.

Table 5: Hypothesis Classification

#	Hypothesis	Main Value
1	The absolute deviation between the predicted and realized number of colli increases throughout the week due to the non-rolling prediction after the last update.	Improve Way of Working
2	The prediction error of the X-1 prediction follows a cyclic pattern.	Data Mining Insight
3	The forecast error is correlated to the forecast volume.	Data Mining Insight
4	There exists a negative linear correlation between the forecast deviation in the non-perishable and perishable A pick zone.	Improve Way of Working
5	The absence of AH employees due to illness follows a seasonal and/or weekly pattern.	Data Mining Insight
6	The average productivity of Flex employees is higher than the productivity of AH employees.	Data Mining Insight
7	The prediction accuracy of the required workforce could be improved by taking the differences in order pick productivity into account between AH and Flex employees.	Improve Way of Working
8	The colli prediction has the largest relative error and is thus the most uncertain variable of the three uncertain variables.	Data Mining Insight
9	The average order picking productivity per shift is negatively linear correlated to the total number of colli to be picked in the entire pick zone per shift.	Data Mining Insight

4. Data Analysis: Testing Hypothesis

This chapter describes the initial data analysis. The first step in this process is to obtain and prepare the data. The available data is described in Section 4.1. In Section 4.2, the results concerning the hypothesis as drawn in the previous chapter are given. Section 4.3 summarizes the main conclusions.

4.1. Available Data and Data Preparation

Albert Heijn stores a lot of data in different types of databases and files. Typically, those databases or files contain a lot of information, or features. Most of those features are most likely not relevant to the analysis that are performed within this thesis. This section describes which features are relevant to the analysis. Accompanying appendices describe the processes required to filter or aggregate the existing data sets such that the required remaining sets are obtained. The available data concerning colli predictions is described in Section 4.1.1, order picker productivity in Section 4.1.2, and absence due to illness in Section 4.1.3.

4.1.1. Colli Predictions

The forecasted and realized colli numbers are stored in a MySQL database, which is updated daily. The forecasts at X-1, X-2 and X-7 are only updated on Wednesdays. The realized demand is updated daily. Appendix C describes how the available data from the MySQL database is transformed, such that all relevant features are obtained. This results in the features as described in Table 6.

Table 6: Features of Colli Prognose Data Set

Feature Name	Description
Date	The production dates.
Shift	Day, evening or night shift of a certain day.
Pick Zone	The pick zone in which the orders where picked. (NP, P-A or P-B.)
Forecast X-7	Colli prognose as given by RE, 7 weeks in advance.
Forecast X-2	Colli prognose as given by RE, 2 weeks in advance.
Forecast X-1	Colli prognose as given by RE, 1 week in advance.
Realisation	Realized number of colli ordered.
Forecast Error / X-1	The percentual difference of X-1 compared to the realized number of colli, see Equation 16.

As can be seen in Table 6 and Equation 16, the percentual difference between the forecast and realization is calculated using the forecasted value of X-1 as the denominator. Thus, for example, in case the forecast was 10 thousand, and the realized 12 thousand, the increase was *confidential* % compared to the forecasted amount. This percentage is also used by the capacity planners to determine if up- or downscaling is required. The formula used to calculate the percentual difference at X-1 is shown in Equation 16.

$$\% \, Difference \; X - 1 \; to \; Forecast = \frac{\# \, Ordered - \# \, Forecasted \; at \; X - 1}{\# \, Forecasted \; at \; X - 1} \tag{16}$$

This same equation could be used for the prediction made at X-2 and X-7. However, the hypothesis tested within this research do not require those values.

4.1.2. Order Picker Productivity and Colli Realizations

Data concerning the order picking process is collected in the Warehouse Management System (WMS). The system only stores the data for one week. Fortunately, Albert Heijn extracts this information and stores the data at their local disk to use for analysis. This data is stored in CSV files, per production day of each DC. Two types of files are stored per day, the first contains the details of entire orders (PBHEAD files), whilst the second contains the details of each order line (PBROW files). Only the second file contains the information on what type of employee is picking the order, an AH or a flex employee. Therefore, the two files must be merged in order to obtain all the required information for the analysis. Appendix D describes how those two files are merged using functionalities of RapidMiner and how the files are filtered and cleaned such that only relevant order lines remain.

Based on this data, three important features can be deducted:

- The realized number of colli picked (per employee type)
- The total picking duration (per employee type)
- The realized order picking productivity (per employee type)

The remaining feature set is shown in Table 7. To determine the number of pickers needed (per type), Equation 17 is used. The aggregate productivity can then be calculated using Equation 18. The productivity per employee type is calculated by using Equation 19.

$$Pickers \; Needed \; Per \; Type = \frac{Picked \; Hours \; Per \; Type}{7{,}5} \tag{17}$$

$$Aggregate \; Productivity = \frac{Total \; Picked \; Quantity}{Total \; Pick \; Duration} \tag{18}$$

$$Productivity \; Per \; Type = \frac{Total \; Picked \; Quantity \; Per \; Type}{Total \; Pick \; Duration \; Per \; Type} \tag{19}$$

Based on the realized number of picked colli, the serverice percentage of can also be calculated. To do so, the realized number of colli ordered from the DC Planning data set is required. Equation 20 can then be used to calculate the realized service percentage.

$$Service\ Percentage = 1 - \frac{Realized\ Colli\ DCPlanning - Picked\ Colli}{Realized\ Colli\ DCPlanning} \tag{20}$$

Table 7: Features based on PBHEAD and PBROW Files

Feature	Description
Date	The production date of the order.
Shift	Day, evening, or night shift of a certain day.
Pick Zone	The pick zone in which the orders where picked.
Picked Quantity FLEX	The number of colli picked by flex employees.
FLEX Picking Hours	The total hours of order picking performed by flex employees.
Flex Pickers Needed	The total number of flex employees needed to fulfil the order required picking hours.
FLEX Productivity	The average productivity of the flex order pickers.
Picked Quantity AH	The number of colli picked by AH employees.
AH Picking Hours	The total hours of order picking performed by AH employees.
AH Pickers Needed	The total number of AH employees needed to fulfil the order required picking hours.
AH Productivity	The average productivity of the AH order pickers.
Picked Quantity Total	The number of colli picked.
Total Picking Hours	The total hours of order picking performed.
Total Pickers Needed	The total number of employees needed to fulfil the order required picking hours.
Aggregate Productivity	The average aggregate productivity of the flex and AH order pickers.
DCPlanning_Realisation	The realized number of colli ordered.
Service_Percentage	The realized service percentage.

4.1.3. AH Absence Due to Illness

The required data concerning absence due to illness of Albert Heijn employees can be retrieved from Interflex. Since Interflex is used as the registration system for all hours of AH employees, this also includes absence due to illness. It is possible to export data from Interflex, however, this results in data lines concerning one clocking of a single employee. The analysis requires the total hours of absence per day. To obtain this information, aggregation of the data set is required. This process is described in Appendix E. The resulting features are aggregated per production day and shift. Those features are shown in Table 8.

Table 8: Features of Absence Due to Illness Data

Feature	Description
Date	Date of the production day.
Pick Zone	The pick zone in which the orders where picked. (NP, P-A or P-B.)
Shift	Day, night, or evening shift.
Total Hours	The total hours that AH employees are scheduled for the order picking task.
Hours Available	The total hours that AH employees are expected to work. (Total Hours minus all fixed absent hours such as holidays and STWA.)
Hours Absent due to Illness	The total hours that AH employees are absent due to illness.
Worked Hours	The total hours that AH employees where available to work. (Hours Available minus the hours Absent due to Illness.)

4.2. Hypothesis Testing

Within this section, a summary of the results of the tested hypotheses as drawn in the context analyse of Chapter 3 are given. To test the hypotheses, the available data as described in the previous subsection is used. In most cases, simple statistical methods are used, which are calculated in Excel files. For some hypotheses, Tableau is used to get a better insight. All those files can be found in the accompanying folder "Hypothesis_Testing", in which the documents are numbered per hypothesis.

A more detailed explanation per hypothesis, including tables and figures, is given in Appendix F.

Hypothesis 1: The absolute deviation between the predicted and realized number of colli increases throughout the week due to the non-rolling prediction after the last update.

To test this hypothesis, the mean absolute deviation, or MAD, to the forecasted number of colli per delivery day is determined based on the available data of 2019 at DCO. Both the non-perishable and perishable pick zone A are analysed. The numerical results are shown in Table 9.

Table 9: MAD per Shift, Pick Zone, and Day, DCO, 2019

Shift	Zone	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Day	NP	*Confidential*						
	PA							
Night	NP							
	PA							

For both pick zones, the hypothesis does not hold. The non-perishable pick zone shows the largest MAD at Thursdays, and a relatively low MAD at Sundays. The perishable pick zone A shows the largest peek at Monday for the Night shift, and Tuesday in case of the day shift.

An explanation for the largest deviation on Thursdays for the non-perishable pick zone could be the usage of box-controlling by the Replenishment department. As decsribed in Section 3.3.1, using box-controlling to reduce the deviation between predicted and realized colli numbers on a certain day, might increase the deviation on subsequent days. Although this might explain the results, we can conclude that the MAD of the forecast does not increase throughout the week due to the non-rolling forecast. And thus, investing in creating a rolling forecast will most likely not be benificial.

Hypothesis 2: The prediction error of the X-1 prediction follows a cyclic pattern.

Although different methods have been applied to identify cyclic patterns in the prediction error at X-1, no clear cycle is identified. As shown in the appendix, the data slightly suggest a weekly or monthly pattern. However, this strongly depends on the period of 2019 that is analysed. Therefore, we cannot conclude that the prediction error made by the replenishment department shows a cyclic pattern.

Hypothesis 3: The forecast error is correlated to the forecast volume.

In case the forecast error is related to the forecast volume, either a positive or negative correlation should be present, when comparing the forecasted colli at X-1 and the percentual deviation of the realization compared to the forecast at X-1.

However, the largest Pearson correlation value, the highest positive or lowest negative value, that is found when analysing the day and night shifts at DCO in 2019, is only 0,276. This correlation is so small that we can conclude that there is no significant positive or negative correlation between the forecasted volume and the error of the forecast.

Hypothesis 4: There exists a negative linear correlation between the forecast deviation in the non-perishable and perishable A pick zone.

In case there exists a negative correlation between the forecast deviation in the non-perishable and perishable A pick zone, Albert Heijn might be able to exploit economies of scale. In case a negative deviation in the non-perishable pick zone is realized, there might be excess employees, which could be transferred to the perishable pick zone, in case a positive deviation is realized in that pick zone.

To determine if a correlation exists, the Pearson correlation coefficient is calculated for the day and night shifts of 2019 at DCO. The correlation values are 0,073 and -0,017 respectively, which are very close to 0 and thus both classified as very weak. Thus, we can conclude that there is no (negative) correlation between the forecast errors in the two pick zones and thus, economies of scale cannot be exploited.

Although it would have been beneficial to exploit economies of scale, in case a negative correlation existed, it is not surprising that this correlation does not exist. In case demand is over- or underestimated it is more likely that the entire customer demand is estimated incorrectly, not only for the perishable or non-perishable pick zone.

Hypothesis 5: The absence of AH employees due to illness follows a seasonal and/or weekly pattern.

Based on autocorrelation and visual analysis of the data of the day shift in the non-perishable pick zone at DCO in 2019, no real seasonal pattern can be identified. However, the results do show peaks in January, February, April, May, June, and November. As shown in Figure 17, those months, at least 40% to 60% of the days illness was registered. Compared to 11% to at most 26% for the other months. In total, only at 34% of the days illness was registered in 2019.

Although there is no clear monthly or seasonal pattern, there is a clear weakly pattern. Within this pattern, Sundays are excluded, since most Sundays no AH employees are working. The pattern can be seen in Figure 18, which shows the percentage of days in 2019 illness was recorded per weekday. The figure shows that on Wednesdays and Saturdays, illness is very low. On almost 16% of the Wednesday's of 2019 illness was registered, and only 4% of the Saturdays recorded illness. Whilst all the other weekdays recorded illness at least on 42% up to 56% of the days. Thus, the absence due to illness follows a three-day pattern, excluding Sundays form this pattern.

Confidential Figure

Figure 17: Percentage of Illness per Month, Non-Perishable, Day Shift, DCO, 2019

Confidential Figure

Figure 18: Percentage of Illness per Weekday, Non-Perishable, Day Shift, DCO, 2019

Hypothesis 6: The average productivity of Flex employees is higher than the productivity of AH employees.

Table 10 shows the mean of the average productivity realized per employee type at DCO in 2019. As can be seen, the difference between the two productivities seems large, indicating a higher productivity reached by flex employees.

To test the significance of this difference, a one-tailed paired t-test is used. Both in case of the day and night shift the test statistic should be greater than 1,65 to conclude that the mean of the distributions is different. As Table 10 shows, both t statistics are high greater than 1,65 and thus resulted in the conclusion that the average productivity of flex employees is significantly higher than the average productivity of AH employees.

Table 10: Sample Mean of Productivity per Employee Type and Shift, Non-Perishable, DCO, 2019

Shift	AH	Flex	Difference	T Statistic
Day	*Confidential*			
Night				

Hypothesis 7: The prediction accuracy of the required workforce could be improved by taking the differences in order pick productivity into account between AH and Flex employees.

To test this hypothesis, both methods are performed based on the data of 2019 and 2020 for the day and night shift. Thus, the required number of flex employees is predicted based on the realized aggregated productivity, and the realized productivity per employee type, as shown in Figure 19 and Figure 20 respectively.

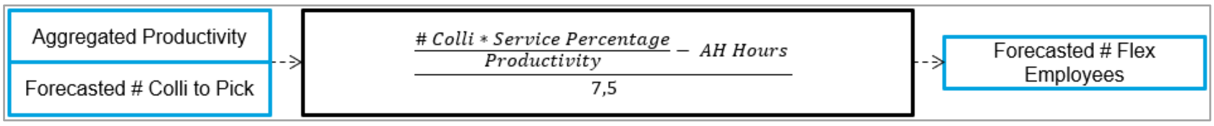

Figure 19: Determining Difference in Case of Aggregated Productivity

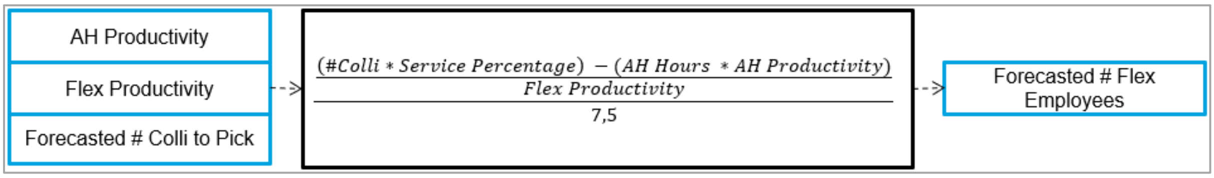

Figure 20: Determining Difference in Case of Separated Productivity

The predicted value of both methods is compared to the realized number of flex employees. The difference between the predicted and realized should be as close to zero as possible. Table 11 shows the average difference realized per prediction method for both the day and night shift. As the table shows, differentiating in productivity reduces the accuracy of the prediction

during the day shift, shows a slightly decrease in the mean difference between the predicted and realized number of employees. However, in the night shift, this is the other way around, thus the current method slightly outperforms differentiating between the two productivities.

Table 11: Mean and Standard Deviation of Difference Between Forecasted and Realized Number of Flex Employees

	Day		Night	
	Difference Aggregated Productivity Figure 19	Difference AH≠Flex Productivity Figure 20	Difference Aggregated Productivity Figure 19	Difference AH≠Flex Productivity Figure 20
Mean	*Confidential*			
Std.Dev.				
T Statistic				
P-Value				

A paired t-test is used to assess whether the differences are significant. Similarly, to the previous hypothesis, the test statistic should be greater than 1,65 to confirm the result of both methods are significantly different. As Table 11 shows, for both shifts, the test statistic is large enough. Thus, we can conclude that differentiating in the productivity between AH and flex employees is less accurate to predict the required number of flex employees.

Hypothesis 8: The colli prediction has the largest relative error and is thus the most uncertain variable of the three uncertain variables.

As shown at Hypothesis 5, the unexpected absence of AH employees due to illness is often zero. In only *confidential* % of the days in 2019 unexpected illness was registered. As shown in Appendix F, assuming that no employees will be absent due to illness will result in an average error of approximately 1 hour is realized. It is not possible to calculate the relative error, since most of the realized values are equal to zero.

However, as Table 12 shows, the relative error can be calculated for the colli prediction and for the aggregate productivity. As the table shows, the average relative error is largest for the colli prediction in both the day and night shift.

Table 12: Average Absolute Relative Error per Uncertain Variable and Shift, Non-Perishable, DCO, 2019

<table>
<tr><th>Error of:</th><th>Day Shift</th><th>Night Shift</th></tr>
<tr><td>Aggregate Productivity</td><td colspan="2" rowspan="2">Confidential</td></tr>
<tr><td>Picked Colli =
X-1 Prediction * Service Percentage</td></tr>
</table>

Hypothesis 9: The average order picking productivity per shift is negatively linear correlated to the total number of colli to be picked in the entire pick zone per shift.

The Pearson correlation coefficient is determined for the day and night shift, and the AH and Flex employee types, resulting in four coefficients. The highest correlation value found is 0,265, which is classified as weak. Since all the correlation values are so close to zero, we can conclude there is no linear correlation between the total colli to pick from the non-perishable pick zone and the productivity of the employees.

However, it might be possible that some other, non-linear relationship exists between the two variables. Due to time limitations, this is not further investigated within this research.

4.3. Conclusion

Within this chapter, the hypotheses as defined in the previous chapter are tested. The hypotheses where classified in two types: hypotheses which could improve the way of working at Albert Heijn, and hypotheses which would give insight in the options for data mining. The hypotheses, their class, and the result as shown in this chapter are summarized in Table 13.

Table 13: Hypothesis Results

#	Hypothesis	Result	Main Value
1	The absolute deviation between the predicted and realized number of colli increases throughout the week due to the non-rolling prediction after the last update.	No	Improve Way of Working
2	The prediction error of the X-1 prediction follows a cyclic pattern.	No	Data Mining Insight
3	The forecast error is correlated to the forecast volume.	No	Data Mining Insight
4	There exists a negative linear correlation between the forecast deviation in the non-perishable and perishable A pick zone.	No	Improve Way of Working
5	The absence of AH employees due to illness follows a seasonal and/or weekly pattern.	Yes, weekly	Data Mining Insight
6	The average productivity of Flex employees is higher than the productivity of AH employees.	Yes	Improve Way of Working
7	The prediction accuracy of the required workforce could be improved by taking the differences in order pick productivity into account between AH and Flex employees.	No	Improve Way of Working
8	The colli prediction has the largest relative error and is thus the most uncertain variable of the three uncertain variables.	Yes	Data Mining Insight
9	The average order picking productivity per shift is negatively linear correlated to the total number of colli to be picked in the entire pick zone per shift.	No	Data Mining Insight

5. Model Options and Experiment Design

In the previous chapter, multiple hypotheses are tested concerning the current method to determine the number of flex employees to request one week in advance of the production day. Based on those results, two different model types are proposed in this chapter, that aim to improve the current prediction method. The methods can only use the data that is available at least one weak in advance of the production day, to align with the agreements of the employment agencies and the update in the WAB.

In Section 5.1, the two model options are described to improve the decision on how many flex employees to request one week in advance of the production day. This includes machine learning to directly predict the number of flex employees and using machine learning to alter the current decision model. Section 5.2 describes the sequence in which the different methods are developed and tested, for which the results are given in Chapter 6. Finally, Section 5.3 describes the available data which can be used as input the machine learning models.

5.1. Two Decision Model Options

Two different model types will be evaluated as model to decide on the number of flex employees to request at the employment agency one week in advance. Both models aim to improve the current decision model, which is shown in Figure 21. The current decision model uses the point forecasts or estimates of the four uncertain variables. By using the equation as shown in the figure, a decision is made on the number of flex employees. This decision does not include a prediction interval, since the equation only uses point forecasts. The two alternative models both allow the usage of prediction intervals. Both models are explained in the following subsections.

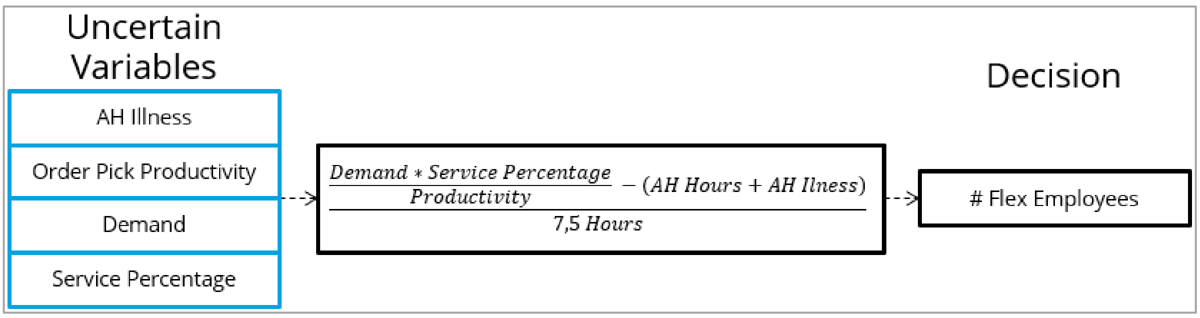

Figure 21: Current Decision Model to Determine the Number of Flex Employees to Request

Directly Predicting the Required Number of Flex Employees

The first proposed method is directly predicting the required number of flex employees using a machine learning algorithm. Thus, the entire equation is replaced by a machine learning algorithm, as shown in Figure 22. Since the equation is replaced by a machine learning model, it is possible to generate prediction intervals, as explained in the experimental design in Section 5.2 and Appendix I.

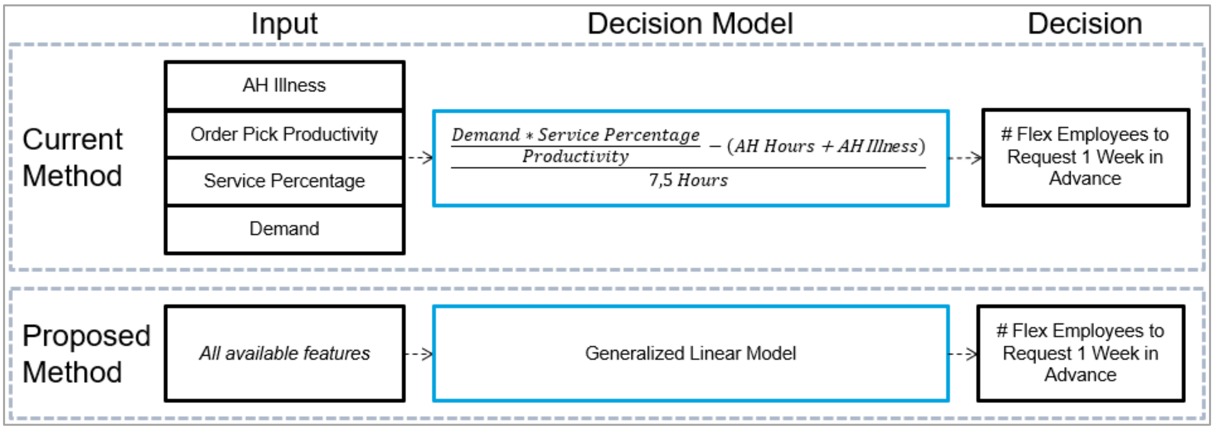

Figure 22: First Decision Model Option: Directly Predicting Required Number of Flex Employees

As shown in Chapter 4, there seem to exist patterns in the input variables when determining the required number of flex employees. For example, the data suggest a weekly pattern in the absence of AH employees due to illness. And there exists a slight weekly or monthly pattern in the prediction error of the replenishment department. Although the patterns are not extremely clear, using machine learning, those patterns might be exploited. Therefore, this first method is tested.

Altering the Current Decision Model

This second model is developed based on the results of Hypothesis 8. In Section 4.2 it is shown by testing this hypothesis, that the number of colli to pick is the uncertain variable with the highest relative error and the greatest impact on the decision. Reducing this error might be possible by applying a machine learning technique to alter the prediction given by the Replenishment department.

Thus, instead of replacing the entire equation, only the order quantity and service percentage are replaced, as shown in Equation 21 and Figure 21, the required number of flex employees is now predicted using Equation 22. The differences in the two equations are highlighted in red.

$$\#Flex = \frac{\frac{Xmin1 * 3\_ServicePercentage}{3\_Aggregate\_Productivity} - Planned\_AH\_Hours}{7{,}5} \qquad (21)$$

$$\#Flex = \frac{\frac{Colli\_Prediction}{3\_Aggregate\_Productivity} - Planned\_AH\_Hours}{7{,}5} \tag{22}$$

Furthermore, generating a prediction interval for the colli prediction, also allows for the generation of a prediction interval of the number of flex employees to request. This second model and the differences to the current model, are depicted in Figure 23.

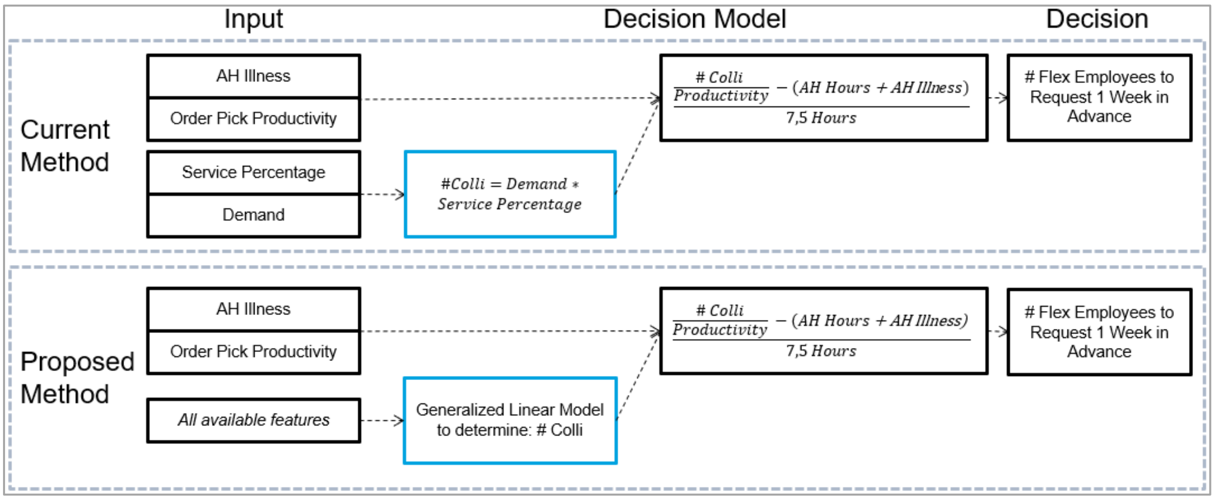

Figure 23: Second Decision Model Option: Altering Colli Prediction in Current Decision Model

5.2. Experimental Design

Within this section, the experiments as used to develop and test the two different models are described. The results of those experiments and the resulting models are given in Chapter 6.

For both proposed models, the same experimental design is used, which is shown in Figure 24. The literature review revealed that most research in similar application areas use neural networks as a machine learning method for workforce predictions. Although this method is most widely used, multiple other methods have been applied and proven to be successful. Therefore, the first step within the model development of both models, is to test multiple machine learning methods. To do so, for both model types, the Auto Model functionality of RapidMiner is used.

Based on the auto modelling results, the best machine learning method is selected for both model types and developed in two custom prediction models in RapidMiner. This is required, since the auto modelling functionally randomly splits the data in a training and test set, whilst we require the data to be split based on the availability of the data. The training data cannot contain data of realizations which lie in the future of the test set. By using a custom model, the data can be split manually. Furthermore, the custom models allow for the determination of prediction intervals.

As shown in Figure 24, the last two steps are evaluating the performance of the both models. And second, performing experiments on the usage of the prediction intervals based on penalty costs. This last step is only performed for the model which resulted in the best performance in step 3. The steps are described in more detail in the following subsections.

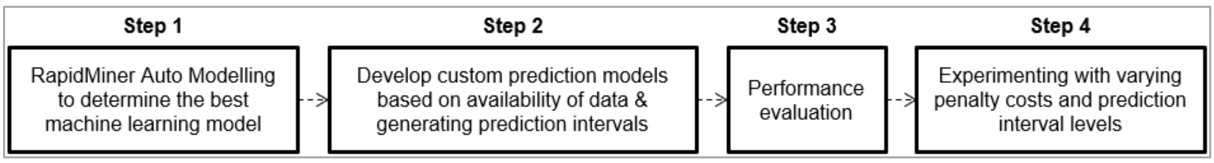

Figure 24: Model Development Process

Step 1: Auto Modelling to Compare the Machine Learning Methods

The auto model functionality of RapidMiner is used to evaluate the performance of four of the proposed machine learning techniques: Generalized Linear Models, Deep Learning, Random Forest, and Gradient Boosted Trees.

By using the auto model functionality of RapidMiner, the parameter settings of the random forest and gradient boosted tree are automatically optimized, as shown in Table 14. However, the parameters of the generalized linear model and the deep learning model are not optimized automatically. RapidMiner did not include this process for deep learning since it would most likely be too time consuming (Mierswa, 2020).

Table 14: Automatically Optimized Parameter Settings

	Random Forest	Gradient Boosted Trees
Number of Trees	Yes	Yes
Maximal Depth	Yes	Yes
Learning Rate	N.A.	Yes

For all models, the Auto Modelling functionality automatically creates features, and selects the features with the greatest weights. The types of features that are engineered are explained in more detail in Section 5.3.

Based on the results of the auto modelling, the best model is selected to predict the required number of flex employees or the number of colli to pick for model 1 and 2 respectively. Those two models are then developed in a custom process in RapidMiner as explained in step 2 a and 2b below.

Step 2 a: Correctly Splitting the Data in Train and Test Set

The first drawback of the auto model functionality of RapidMiner is the fact that the data is split randomly into a training and test of 60% and 40% of the data set respectively. The split is performed by shuffled sampling. This means that the subsets are chosen by randomly selecting examples. Thus, the training and test set consist of dates which are randomly picked from 2019 and 2020. However, in the real world, it is not possible to train the prediction model this way. The predictions can only be made based on the data of the dates that are already in the past. Therefore, the auto model results only give an indication of the performance of the models.

In a custom RapidMiner process the data can be split manually, allowing a split which does take the time series in mind. The data can be split into a training set consisting of 60% of the data, which is simply the first 60% of the observations. The test set then contains of the remaining 40% of the data, of which the observations are all made after the observations of the training set.

Step 2 b: Generating Prediction Intervals

The second drawback of the auto model process is the fact that it is not able to generate prediction intervals. In Section 2.2.2 of the literature review, the relevance of prediction intervals is described and the equations are given to calculate those prediction intervals. In Appendix H, the custom RapidMiner process is shown that uses Equation 1, 2, and 3 from the literature review to generate the prediction intervals.

Step 3: Measuring Performance

As described in the literature review, the mean average percentage error (MAPE), is an appropriate method to compare different prediction methods. Thus, this method is applied when comparing the performance. However, this only provides an insight in the better prediction method, not in the improvement of the performance for Albert Heijn.

It is not possible to relate the prediction error directly to costs. For example, in case the number of required flex employees is underestimated, this does not mean that those hours will not be fulfilled by last minute requests. So, it is not possible to translate those hours to overworking costs.

In case of overestimation of the required number of flex employees, Albert Heijn has arranged that the employment agency is responsible for the possible cancellation costs. Only in case the employees have already arrived at the distribution centre, Albert Heijn is forced to pay them at least 4 hours. In case this happens, the shift leaders always try Albert Heijn order pickers first, if they want a day off. If there are enough Albert Heijn employees that are willing to take a day of, there is no need to cancel working hours of flex employees, not resulting in any additional costs.

However, as described in the problem statement, the tariffs of the employment agencies are based on certain rules. This includes the rule that up- or downscaling the requested number of employees with *confidential* % is allowed. This rule is included in the tariffs the employment agencies handle. In case this *confidential* % marge can be reduced, the tariffs will most likely reduce as well. Thus, by translating the performance of the prediction to a relative error, relative to the predicted value, this performance can be measures. Equation 23 shows how this KPI value can be calculated.

$$KPI_{Relative\_Error} = \sum_{i \in I} \frac{|Prediction_i - Realisation_i|}{Prediction_i} * 100\% \qquad (23)$$

Step 4: Making Decisions based on the Prediction Intervals

Although it is not possible to relate the over- or underestimation of the required number of employees to the costs, it is possible to evaluate the results assuming certain over- or underestimation “costs”. To do so, we use the KPI “Penalty”. Instead of assigning costs to under- or overestimation, a penalty value is assigned to the error. By varying the “penalty costs” in case of under- or overestimation, it is possible to get an insight in the performance of the predictions.

Equation 24 shows how the penalty of a prediction is determined. Based on all predictions made by a model for the test set I, the total KPI value is then calculated using Equation 25.

$$Penalty = \begin{cases} Penalty_{Underestimation} * |Error| & In\ case\ Underestimated \\ Penalty_{Overestimation} * |Error| & In\ case\ Overestimated \end{cases} \quad (24)$$

$$KPI_{Penalty} = \sum_{i \in I} Penalty_i \quad (25)$$

Based on the results of the previous steps, the best model will be selected to perform this last experiment with. In this experiment, five penalty combinations are evaluated. The first is simply equal penalties for both under- and overestimation. The other experiments assume a higher penalty for either under- or overestimation. First only making one of the errors twice as costly, and second making the error four times as costly. The penalty costs per experiment are shown in Table 15.

Table 15: Experimental Design Factors Under- and Overestimation

Experiment Number	Factor Underestimation	Factor Overestimation
1.	1	1
2.	2	1
3.	1	2
4.	4	1
5.	1	4

In addition to only determining the penalty for each experiment based on the given predicted value, the experiments are extended to determine the best level in the prediction interval for each penalty assignment. To do so, the prediction is altered, such that it lies somewhere in the prediction interval. An example is shown in Figure 25. This could be an example of an experiment in which overestimation is more costly than underestimation. If that is the case, it is preferable to make an underestimation and thus use a prediction somewhere in the prediction interval, lower than the original prediction.

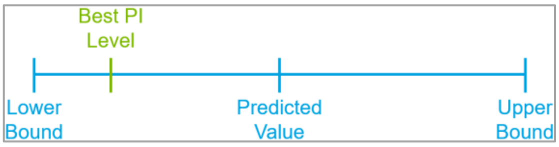

Figure 25: Example Best Prediction Interval Level

To determine the best prediction level, the Excel Solver tool is used. The total KPI Penalty value is minimized, by altering the prediction interval level which is used. The updated prediction is calculated by using Equation 26.

$$Updated\ Prediction = Lower\ Bound + Best\ PI\% * PI_{Width} \quad (26)$$

5.3. Data Sets and Feature Generation

The available data is described in Section 4.1 and the preparation of the data in Appendices C, D, and E. To use this data for the models in this chapter, the data of the three different data sources are aggregated. This process is described in Appendix F. The aggregation of datasets results in the features as shown in Table 16 on the next page.

The table shows that the averages of the previous three weeks of the realized number of employees and productivity per employee type are generated. However, before those averages are taken, outliers are removed from the initial dataset. This is done for the features: Realized AH, Flex, and Aggregate Productivity, Realized Flex Hours Needed, and Realized Service Percentage. As explained in Appendix F, the data of those features showed some serious outliers which are unexplainable or unrealistic. After the outliers are removed, the averages over the realisations of the past three weeks are calculated, resulting in the features starting with "3_". This process is also described in Appendix F.

The data set in Table 16 also shows the values that the current method of the capacity planners would result in, as also shown in Equation 27. In Chapter 4 the hypothesis was tested if differentiating between the AH and Flex productivity would be beneficial to alter the current decision method. However, this hypothesis did not hold, and thus the two proposed models are only compared to the original current prediction method using the aggregate productivity, for which the calculation is shown in Equation 27.

$$Pred\_AGG = \frac{Repl\_Colli\_xmin1 \; * 3\_ServicePercentage}{3\_AGG\_Prod} - Planned\_AH \tag{27}$$

Table 16: Available Data Set

Category	Feature	Description
Basics	Date	Date of the production day.
	Shift	The shift, either day or night.
Planned	Planned_AH	The hours of AH employees scheduled on the order picking task.
Predicted	Repl_Colli_ xmin1	Predicted number of colli to pick given by the replenishment department one week in advance of the production day.
	Repl_ Colli_xmin2	Predicted number of colli to pick given by the replenishment department two weeks in advance of the production day.
	Repl_ Colli_xmin7	Predicted number of colli to pick given by the replenishment department seven weeks in advance of the production day.
	Pred_AGG	The prediction of the required number of flex employees based on the current method. (Equation 27)
Realized	R_Colli	Realized number of colli ordered.
	R_Service_ Percentage	Realized service percentage.
	R_AH_Hours	Realized hours AH employees performed the order picking task.
	R_FLEX_Hours	Realized hours flex employees performed the order picking task.
	R_AH_Prod	Realized productivity of AH employees on the order picking task.
	R_FLEX_Prod	Realized productivity of flex employees on the order picking task.
	R_AGG_Prod	Realized aggregate productivity of AH and flex employees on the order picking task.
Average of Realized Previous 3 Weeks	3_Colli	Average realized number of colli ordered of the previous 3 weeks.
	3_Service_ Percentage	Average realized service percentage of the previous 3 weeks.
	3_AH_Hours	Average realized hours AH employees performed the order picking task of the previous 3 weeks.
	3_FLEX_Hours	Average realized hours FLEX employees performed the order picking task of the previous 3 weeks.
	3_AH_Prod	Average realized AH productivity of the previous 3 weeks.
	3_FLEX_Prod	Average realized FLEX productivity of the previous 3 weeks.
	3_AGG_Prod	Average realized aggregate productivity of the previous 3 weeks.

Since Table 16 is very large, Table 17 gives an overview of the relations per feature. As this table shows, the features starting with “3_” are the averages taken from the realized values from the same weekday the previous three weeks. Those features can be used as input to the machine learning models, since this data is available at least one weak in advance. Taking the average of the last three weeks is also in line with the current prediction method of the capacity planners.

All the realized values are shown in the middle column and cannot be used as input to the machine learning models. The two features highlighted in green “R_FLEX_Hours” and “R_Colli” are the features that the machine learning model must predict in model 1 and 2 respectively.

The first column can be used as input to the machine learning models and contains the predicted values, either by the replenishment department in case of the number of colli to pick, or by the capacity planner in case of the planned AH hours and predicted number of flex employees by the current method.

Table 17: Overview Features: Predictions – Realizations – Averages

Predicted	Realized	Average Realized Last 3 Weeks
Repl_Colli_ xmin1 **Repl_ Colli_xmin2** **Repl_ Colli_xmin7**	R_Colli	3_Colli
	R_Service_Percentage	3_Service_Percentage
Planned_AH	R_AH_Hours	3_AH_Hours
Pred_AGG	R_FLEX_Hours	3_FLEX_Hours
	R_AH_Prod	3_AH_Prod
	R_FLEX_Prod	3_FLEX_Prod
	R_AGG_Prod	3_AGG_Prod

The features such as shown in Table 16 and Table 17 are not directly usable by all machine learning algorithms. For example, the feature "Date" is not in a form such that a machine learning algorithm is able to discover patterns. Therefore, these features should be transformed before feeding the data into the machine learning algorithms.

When using the auto model functionality of RapidMiner, the features are automatically transformed such that the data makes sense to the models. For example, the feature "Date" is transformed into multiple binominal variables such as "Day_of_week=1", which is 1 if the date is a Monday, and 0 otherwise. Similarly, features are generated for months or periods such as "Month=11", or "Quarter_of_year=3". Furthermore, the feature "Shift" is currently stored as a text feature which is either "Day" or "Night". This feature could be changed into binominal features as well, e.g. "Shift=Day" and "Shift=Night".

In case of custom RapidMiner processes, the features must be altered manually such that the machine learning algorithms are able to handle all values.

Based on the available data, and after the outlier removal as shown in Appendix F, Table 18, shows the resulting data set consists of 470 records. The number of day and night shifts is approximately equal. Splitting the data in a training and test set of 60% and 40% respectively, results in data sets of 282 and 188 records.

Table 18: Number of Records per Data Set

Data Set	Number of Records	Records Day Shift	Records Night Shift
Entire Data Set	470	264	206
Train Set = 60%	282	153	129
Test Set = 40%	188	111	77

6. Experiment Results

In the previous chapter two model types are proposed and the experiments to develop those methods are explained. Within this chapter the most important results are shown. In Section 6.1 the results of directly predicting the required number of flex employees by using machine learning techniques are given. Section 6.2 shows the results of altering the colli predictions by machine learning to improve the performance of the decision model which is currently used. The chapter is concluded in Section 0.

6.1. Predict Number of Flex Employees

Appendix G shows the results of the first step in which the Auto Modelling tool of RapidMiner is used. Based on those results, the Generalized Linear Model is selected as the best method to predict the required number of flex employees. Therefore, this method is developed further in a custom process. This process is also described in Appendix G. This process ensures that the model is trained based on the first 60% of the data, and the performance is evaluated based on 40% of the data, which is all of dates after the dates in the training set.

Selected Features

In the Auto Modelling of RapidMiner, a subset of the features is selected to use for the Generalized Linear Model. This subset is shown in Table 19 with the coefficients in case of the Auto Model functionality and in the custom model. Since the custom model uses a different training set, the coefficients per feature vary per model. Additionally, a custom model is developed that is allowed to select all the available features to determine the prediction. This resulted in the features and coefficients as shown in Table 20.

Table 19: Coefficients Generalized Linear Model Features

Feature	Auto Model	Custom Model
Date:half year = 1	-2,43	-0,582
Date:day_of_week = 5	2,77	1,235
Date:day_of_month	0,09	
Date:year	3,14	
Total_Planned_Hours	-0,07	-0,004
DCP_xmin1	0,00	0,000
3_FLEX_Hours	0,22	0,267
3_Service_Percentage	11,08	-1,657
Intercept	-6.348,35	30,300

Table 20 shows, the day of week is an important feature in determining the required number of flex employees, since all of those features, except for the third weekday, are included in the model. This makes sense, since the production volumes vary significantly based on the weekday. Furthermore, the forecast error on the predicted number of colli, and the absence of Albert Heijn employees also suggest weekly cyclic patterns, as shown in hypothesis 1, 3, and 5.

Furthermore, the model does include the averages of the realized flex hours and productivity, and AH and aggregate productivity. As well as the predictions given by the replenishment department at 7, 2 and 1 week in advance. This is in line with the simple calculations of the capacity planner.

Table 20: Coefficients Generalized Linear Model Features Custom Model

Feature	Coefficient
Date:day_of_week = 5	0,742
Date:day_of_week = 6	0,560
3_FLEX_Hours	0,173
Date:day_of_week = 4	0,097
DCP_xmin1	0,000
DCP_xmin2	0,000
3_Colli	0,000
DCP_xmin7	0,000
ACKSHIPDATE	0,000
days_diff(Date, Today)	0,000
Total_Planned_Hours	-0,006
3_AH_Prod	-0,013
3_FLEX_Prod	-0,024
3_AGG_Prod	-0,025
Date:quarter = 1	-0,351
Date:day_of_week = 1	-0,354
Date:day_of_week = 7	-0,694
Date:day_of_week = 2	-4,045
Intercept	-106,142

Prediction Accuracy

The prediction accuracy is determined for the current method, the GLM method whilst using the features selected by Auto Modelling, and the GLM method using all the available features. The results in terms of the MAD and MAPE are shown in Table 21. The table also shows the prediction interval width of the models. In the current situation, it is not possible to generate prediction intervals, thus the PI width is not available. As the table shows, the model which can select all the features performs best, with a MAPE of *confidential* % compared to *confidential* % when using the current prediction method.

The table also shows that the method with all features also outperforms the model only using the Auto Model features and is therefore also able to create a smaller prediction interval. However, for both models, the prediction interval widths are quite large. On average, the predicted required number of flex employees is about 54. With an interval width of *confidential*, this means that the actual required number of employees is with 95% certainty in the range of *confidential* employees.

Table 21: Results: Predict #Flex Employees

	Current	GLM AM Features	GLM All Features
MAD			
MAPE		*confidential*	
Prediction Interval Width			

Figure 26 visualizes the differences between the current decision model and the proposed model. The figure also includes the performance of both models in terms of the MAPE. The best model is the proposed model, this is highlighted in green.

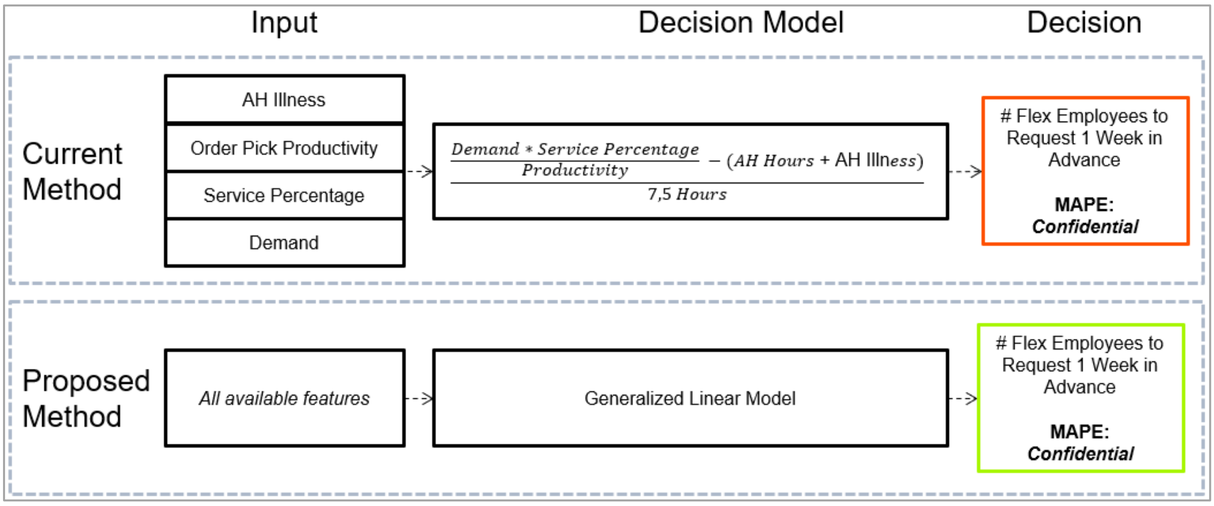

Figure 26: Current and Proposed Model Directly Predicting Flex Including Results

As described in Section 5.2, the most important measure is the error of the prediction, relative to the predicted value. Based on the request made one week in advance, the number of employees can be up- or downscaled *confidential* %. Therefore, the KPI value as calculated by using Equation 28 gives a good indication of the average absolute error, relative to the predicted value.

$$KPI = \sum_{i \in I} \frac{|Prediction_i - Realisation_i|}{Prediction_i} * 100\% \tag{28}$$

Table 22 shows the KPI value for the three prediction methods: current, GLM using the Auto Model features, and GLM able to use all features. As the table shows, the last model is again the best performing model. The table additionally shows the percentage of time, the relative error was below *confidential*%, *confidential*%, and *confidential*%. In the best-case scenario, all predictions should be below *confidential* %. However, the table shows that in the current situation, only *confidential* % of the predictions achieves this score. The GLM models slightly improve this score, achieving *confidential* % and *confidential* % of the predictions within a *confidential* % range.

Table 22: KPI Performance Directly Predicting #Flex

	Current	GLM AM Features	GLM All Features
Average Predicted #Flex Employees			
KPI: Relative Error		*Confidential*	

It is apparent that the current method, only stays within a *confidential* % range of the predicted value in *confidential* % of the observations. This might be explained by the fact that the request made by the capacity planner includes more than just the non-perishable colli pick zone. This increases the total number of requested flex employees, also making the *confidential* % range larger. The average number of requested flex employees is also shown in Table 22, which also shows that the best performing method in terms of the KPI, is also the method with the highest average prediction value, making the *confidential* % range larger. For example, in case of the current method, *confidential* % is on average equal to *confidential* employees, whilst with the best performing GLM method *confidential* % is on average equal to *confidential* employees.

Although this nuance in performance in terms of the KPI is important, the GLM method still outperforms the current method, since the MAPE is lower.

6.2. Altering Colli Predictions

Since Hypothesis 8 showed that the colli predictions made by the replenishment department, combined with the predicted service percentage, has the largest relative error of the uncertain variables, this method to determine the number of colli to pick is evaluated. By using machine learning algorithms an altered prediction of the number of colli to pick is made, which can then be used in the current decision model with the simple equation.

The results of the auto modelling, step 1, can be found in Appendix H. Again, the Generalized Linear Model performed best in this step, resulting in the custom development of a GLM, using the correctly split train and test set. The features used in the Auto Model and custom model are compared below, followed by a comparison of the prediction accuracy.

Selected Features

Table 23 shows the selected features and their coefficient in the Auto Model experiment. Additionally, the table shows the coefficients of those features in when using a custom model with the correctly split data.

The first thing that is apparent is the fact that this model does include the different shift types as features, whilst when predicting the required number of flex employees, the shift type did not result as one of the features. However, Table 24, shows the features selected by the custom model, in which all features are allowed, and this model did not include the shift features.

Table 23: Coefficients Generalized Linear Model Features

Feature	Auto Model	Custom Model
Shift.DAY	-1451,19	-1202,76
Shift.NIGHT	1822,69	1426,53
Date:quarter = 4	5322,85	
Date:half year = 2	-1625,06	-2577,39
Date:day_of_week = 4	-2627,61	1659,80
Date:month_of_quarter = 1	1562,59	1422,00
Date:day_of_month	93,62	
DCP_xmin1	0,27	0,32
DCP_xmin7	0,08	
3_Service_Percentage	14192,43	16827,72
sqrt([DCP_xmin1])	276,85	287,17
Intercept	-43783,28	-45925,79

Again, the results show, especially Table 24, that the day of the week is of great influence and therefore selected as feature. Additionally, the moth of the quarter seems of influence too, appearing both in the auto model features and selected by the custom model. Furthermore, the prediction given by the replenishment department and the average realized service percentage are included in all models as well. Which makes sense, since those two features form the current prediction method. It makes less sense that the average AH productivity is included in the model.

Table 24: Coefficients Custom GLM Features

Feature	Coefficient
Date:day_of_week = 7	-6575,4
Date:day_of_week = 2	-3784,9
Date:day_of_week = 3	-1019,5
Date:month_of_quarter = 2	-854,5
3_AH_Prod	-3,7
Date:quarter = 3	-2,5
DCP_xmin1	0,763
Total_Planned_Hours	2,341
Date:month_of_quarter = 1	665,711
Date:half_year = 1	946,564
Date:day_of_week = 4	1084,013
Date:day_of_week = 5	1110,039
Date:quarter = 2	1873,279
3_Service_Percentage	15031,061

Prediction Accuracy

In this model type, the machine learning method is used to determine the required number of colli to pick. This required number of colli to pick is compared in terms of the MAPE, to the X-1 forecast given by the Replenishment department, multiplied by the expected service percentage, as shown in Equation 29.

$$Current\ Colli\ Prediction = Xmin1 * 3\_ServicePercentage \tag{29}$$

As Table 25 shows, the MAPE of the GLM models does outperform the predictions based on the average service percentage and the X-1 prediction of the Replenishment department. However, although the prediction in the number of colli to be picked does improve, the MAPE of the prediction on the required number of flex employees did not improve. This is also visualized in Figure 27 in which the MAPE for the colli prediction and flex prediction is shown for both models. Per step, the best MAPE is highlighted in green, whilst the worst is highlighted in red.

Based on those results, we can conclude that the current prediction method is performing better, in case the colli predictions are less accurate. Table 26 shows that the Generalized Linear Models tend to underestimate the number of colli to pick more often, which is might cause the difference in the performance of the flex prediction.

Table 25: MAPE Prediction #Colli, then #Flex

Features	Colli		# Flex	
	GLM	Current	GLM	Current
Auto Model Features	*Confidential*			
All Features				

Table 26: Over-/Underestimation Predicting #Colli to Pick

	Current	AM Features	All Features
Underestimated	*Confidential*		
Overestimated			

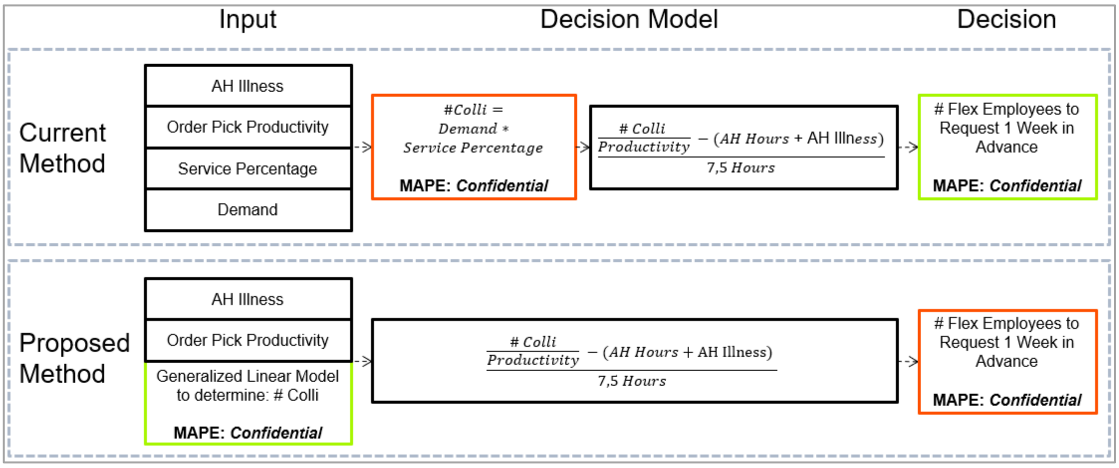

Figure 27: Current and Proposed Model Predicting #Colli, then Flex, Including Results

Figure 27 clearly shows the improvement in the input variables for the decision model. However, the performance of the output of the decision is less accurate. This indicates that there must exist some relations between the four input variables, which cancel each other's errors out.

As Table 27 shows, the GLM models thus also perform worse concerning the KPI and the *confidential* % up- or downscaling performance than the current method. The number of times the actual required number of flex employees is within a *confidential* % range of the prediction, decreases with approximately 3%. As already explained when analysing the results of directly predicting the number of flex employees, the *confidential* % interval decreases, in case the predicted number of employees is smaller. As Table 27 shows, the average predicted value is smaller for both GLM models, compared to the current method, which might be an explanation why the GLM models perform worse.

Table 27: KPI Performance Predicting #Colli then #Flex

	Current	GLM AM Features	GLM All Features
Average Predicted #Flex Employees	*Confidential*		
KPI: Relative Error			

6.3. Penalty and Prediction Interval Evaluation

The final step in the model generation and evaluation process, is the evaluation of the performance based on penalties for under- or overestimation and the optimization of the prediction based on the prediction interval and penalty costs. Since directly predicting the required number of flex employees resulted in the smallest error, this method is selected to perform the experiments on the prediction intervals and related penalties.

Performance based on Current Predictions

Table 28 shows the average total penalties per experiment. As the table shows, the average penalties increase in case the factors get higher. To make those values comparable, the averages should be divided by the sum of the penalty factors, as shown in Equation 30. This weighted penalties per experiment are also shown in Table 28.

$$Weighted\ Penalty = \frac{Penalty * |Error|}{Factor_{Under} + Factor_{Over}} \tag{30}$$

As the table shows, increasing the factor of underestimation (experiment 2 and 4) results in a lower average weighted penalty. Whilst increasing the penalty costs for overestimation results in higher weighted penalty costs. This can be explained by the fact that the with the current predicted values, the number of flex employees is underestimated in 45% of the cases and overestimated in 55% of the cases, as also shown in Table 29. Therefore, assigning more costs to underestimation, only results in an increased penalty in 45% of the cases, whilst overestimation results in 55% predictions which become “more expensive”.

Table 28: Experiment Results Penalties

Experiment Number	Factor Underestimation	Factor Overestimation	Average Penalty	Weighted Penalty
1.	1	1		
2.	2	1		
3.	1	2	*Confidential*	
4.	4	1		
5.	1	4		

Table 29: Percentage of Predictions Under- or Overestimated

	Underestimated	Overestimated
Required Number of Flex Employees	45,2%	54,8%

Determining Best PI Level and Performance

Using the prediction intervals, the performance of the predictions can be improved. As shown in the previous subsection, the current predictions overestimate the required number of employees more often than underestimating. This makes the current prediction less favourable in case overestimation is more costly. Using the Excel Solver tool, it is possible to determine which level in the prediction interval should be selected, such that the total penalty costs are minimized.

The current weighted penalty and the weighted penalties based on the optimal PI levels are shown in Table 30. The table also shows the prediction interval level that is required. As the table shows, the best PI level increases in case underestimation is more costly and decreases in case overestimation is more costly. This makes sense, since altering the predicted value to a higher value, decreases the chance of underestimation, and vice versa. This is also shown in Figure 28 and Figure 29.

Table 30: Experiment Results, Optimizing PI Levels

Exp	Factor Under-estimation	Factor Over-estimation	Weighted Penalty Forecasted Value	Weighted Penalty Optimized PI	Optimized PI Level
1.	1	1	*Confidential*		
2.	2	1			
3.	1	2			
4.	4	1			
5.	1	4			

Confidential Figure

Figure 28: Results Increasing Penalty for Underestimation

Confidential Figure

Figure 29: Results Increasing Penalty for Overestimation

Table 31 shows the percentage of times the updated prediction value under- or overestimates the realized required number. As the table shows, in case of equal penalty costs, this ratio is 50-50%, which makes sense. Making underestimation more costly (experiment 2 and 4), results in lower percentages of times the number is underestimated. And vice versa for the overestimation experiments (3 and 5).

Table 31: Percentage Under- and Overestimated based on Updated Predictions

Exp. Nr.	Underestimated	Overestimated
1.	*Confidential*	
2.		
3.		
4.		
5.		

6.4. Conclusion

Within this chapter, a summary of the results of the two different predictions methods are given. Table 32 shows the summary of the performance of the current method, the best performing machine learning method to predict the required number of flex employees directly, and the best performing method to alter the colli prediction and then use the current decision method.

As the table shows, altering the colli prediction by using machine learning is not outperforming the current prediction method to determine the required number of flex employees. However, as shown in Section 6.2, the prediction on the number of colli to pick is improved. Based on those results, we can conclude that there must exist some relation between the four input variables, which cancels each other's errors out. Therefore, improving only the colli prediction results in an overall worse result.

However, directly predicting the required number of flex employees by using a Generalized Linear Model, does outperform the current prediction method. This means that the machine learning model is able to include those relations in the uncertain variables in the model. The most important features in determining the required number of flex employees are the day of week, the predicted values (X-7, 2, and 1) from the Replenishment department, and the averages of the realized productivity the three previous weeks.

By using the GLM to predict the required number of flex employees for the non-perishable regular pick zone, the predicted number of flex employees is *confidential* % more often within a *confidential* % range of the actual required number of flex employees compared to the current method. This improves the position of Albert Heijn in the negotiations with the employment agencies.

Table 32: Overview Results Prediction Methods

	Current	Predict Flex Directly	Predict Colli then Flex
MAD			
MAPE			
PI Width			
KPI: Relative Error		*Confidential*	

In addition to the development and evaluation of the two prediction models, the model to directly predict the required number of flex employees is analysed in more detail. With five experiments it is shown that it is possible to improve the performance of the model by altering the prediction to another value within the prediction interval. In case of equal costs for under- or overestimation, the best prediction interval level is 47%, since the current model slightly tends to overestimate the required number of employees.

7. Conclusions and Recommendations

This chapter concludes this thesis by addressing the main conclusions and recommendations. The conclusions are given in Section 7.1. The main limitations of this research are described in Section 7.2. Based on the conclusions and limitations, this thesis is concluded with two types of recommendations in Section 7.3. First, recommendations on the way of working at Albert Heijn are given, and second recommendations on future research are given.

7.1. Conclusion

The goal of this research was to improve the prediction of the required number of flex employees, based on data that is available at least one week in advance.

To do so, relevant literature was reviewed and in the context analysis the current processes where described in more detail. The context analysis consisted of multiple interviews with Albert Heijn employees who are currently responsible for the workforce management process. Based on the literature review and the context analysis, multiple hypotheses were drawn concerning the workforce management process of the flex employees in the DCs of Albert Heijn. The main values of those hypotheses were classified in two categories, identifying characteristics that could be exploited when using data mining to predict the required number of flex employees, or improving the current way of predicting by simpler adjustments.

Based on those hypotheses, we conclude that some assumptions of the interviewees of the context analysis, are not true. For example, the usage of a rolling forecast would not improve the prediction accuracy of the required number of flex employees. Additionally, the performance of the current prediction method could not be improved by differentiating between AH and Flex productivity, although the productivity differs significantly per employee type.

In Chapter 5, two alternative prediction methods are proposed to predict the required number of flex employees. As shown in Chapter 6, one of those methods is able to outperform the current prediction method, whilst the other method does not outperform the current method.

By using a Generalized Linear Model to directly predict the required number of flex employees for the regular non-perishable pick zone, the MAPE of the prediction could be reduced from *confidential* % to *confidential* %. More importantly for the DCs, is the error relative to the predicted value. The average of this error reduces from *confidential* % to only *confidential* %.

The Generalized Linear Model was developed using RapidMiner. This allowed for automatic feature generation and selection for the models. Among the selected features for the models where, the shift and day of week, the predictions given by the replenishment department, and the average realizations of the previous three weeks.

Since this research has focussed only on the regular, non-perishable pick zone, the results are expected to improve even further in case more pick zones are included in the prediction. The total requested number of flex employees will increase in this situation, which also makes the *confidential* % range in which the prediction should lay to satisfy the employment agencies, even larger.

Based on the promising results, it is recommended to Albert Heijn to continue the research into the usage of Generalized Linear Models to predict the required number of flex employees. Suggestions on further research are given in the following section.

In addition to the improved predictions, using a Generalized Linear Model also reduces a lot of manual work. Since each distribution centre has their own capacity planner, the automation reduces the required work of five employees. Additionally, the automation reduces the possibility of human errors in the calculations.

7.2. Limitations

The research at hand is subject to several limitations. Some of those limitations are the result of the limited time, whilst others are caused by the limited availability of data.

Focus on Non-Perishable Colli Circuit

As described in the introduction, the research is narrowed down to the non-perishable pick zone of the distribution centre in Zwolle (DCO). More specifically, the machine learning models are only developed for the regular colli pick circuit of the non-perishable pick zone. This was selected since this circuit and pick zone represent the majority of the colli to be picked. However, the other pick zone and circuits require order picking employees as well.

Handling of Break Times

The PBHEAD and PBROW files are used to obtain the total order picking duration on a certain production day. However, those files do not store information on the breaks of employees. It might be possible that an employee took a break during the order picking of a specific order, increasing the total pick duration of the order, whilst the employee was not performing the picking task. Not taking this into account might result in a longer picking duration than was actually required. This also influences the realized productivity.

In future research, productivity could be determined more accurately by taking those breaks into account. If the information is not collected by Albert Heijn, it might be possible to detect orders including breaks in the PBHEAD and PBROW files. For example, by searching for orders of which the pick times of two consecutive picks is significant, for example more than 10 minutes. However, this might have also been caused by other problems, such as spillage, or waiting for pick replenishment.

Limitation in Available Data

There are two types of limitations in the available data. First, the available data is limited, resulting in small training and test sets for the machine learning models. However, it is uncertain if a larger data set, containing information of dates further back in time will actually improve the predictions. The distribution centres are constantly trying to improve their ways of working. In some cases, this has a great impact on the required number of employees. For example, some distribution centres have been remodelled to reduce the travelling distances of the order pickers, reducing the required number of employees. Similar projects happen over time, thus it is uncertain whether more historic data will actually improve the predictions.

Second, the prediction models developed within this thesis only use a limited number of features to train the model. However, there might be additional features that could improve the predictions. For example:

- The number of stores, classified per size, delivered from the distribution centre. Since this is assumed to be correlated to the total number of colli to be picked.
- Weather forecasts, since the weather can influence customer behaviour and thus colli realisations. Additionally, the order picking productivity is correlated to the weather, in case it is hot outside, it will be hot in the distribution centres and the employees will work slower.
- The number of new flex employees. It is known that flex employees only stay for a limited amount of time. However, when new flex employees are hired, those employees must be trained first, resulting in a lower productivity. In case there are a lot of employees leaving and thus a lot of new employees, this would increase the required number of employees.

7.3. Recommendations

The recommendations are subdivided into two subsections. First, the recommendations to improve the current way of working are addressed. Second, the recommendations for further research are given.

Improve Current way of Working

Although some of the hypotheses drawn to improve the current way of working where not true, they did result in useful insights. For example, based on the result, we advise not to invest in creating a rolling prediction method for the colli predictions. Although multiple capacity planners assume this might improve the prediction accuracy, this hypothesis does not hold.

During the interviews as part of the context analysis it also became clear that all capacity planners use their own methods. Although the data used in this research was only from DCO, the distribution centre in Zwolle, it is advised to implement the same method at each distribution centre. For example, the distribution centre in Zwolle did not plan the days off of the AH employees, whilst the distribution centre in Peijnacker makes clear yearly schedules, reducing the uncertainty in the available number AH employees.

Furthermore, it would be beneficial if the decisions could be automized. During multiple interviews it became clear that the process is time consuming and takes a lot of manual work and adjustments. This is not only costly in terms of employee costs, but also prone to errors. Since the process consists of simple calculations, those calculations could be automized. The expertise of the capacity planners is most likely still required, to detect anomalies and handle exceptional days, such as holidays with different shifts etcetera.

Future Research

Since the Generalized Linear Model showed promising results, it is recommended to further research the possibilities of predicting the required number of flex employees by means of machine learning. Possible directions for further research are summed up below.

The current research is limited to the regular, non-perishable colli pick zone. In future research, all pick zones should be included to make it possible to apply the method to the entire prediction process.

Furthermore, within the current research, features are selected based on the performance of the RapidMiner tool. However, there exist multiple methods to perform feature selection. In the limitation section, some possible additional features are already given. By means of feature selection, the performance of the model might improve.

Finally, the impact of special holidays is not taken into account in this research. Although outliers are removed where applicable, this does not mean that all holidays are classified as outlier. A quite straight forward outlier is Christmas, which is always on the same date, and thus could easily be removed from a data set. However, other holidays such as Easter, Whit Sunday, Ramadan, etcetera, all change per year. The predictions around those holidays are often harder to make, and therefore most likely result in greater errors. Additional research concerning those special days could improve the results of the model used for the standard days.

8. References

Ahold Delhaize. (2019, July 26). *z.* Retrieved from Heritage: https://www.aholddelhaize.com/en/about-us/company-overview/heritage/

Bhulai, S., Koole, G., & Pot, A. (2008). Simple Methods for Shift Scheduling in Multiskill Call Centers. *Manufacturing & Service Operations Management, 10*(3), 411-420. doi:https://doi.org/10.1287/msom.1070.0172

Bou-Hamad, I., & Jamali, I. (2020). Forecasting Financial Time-Series Using Data Mining Models: A Simulation Study. *Research in International Business and Finance, 51*. doi:https://doi.org/10.1016/j.ribaf.2019.101072

de Lang, M. (2019). *A Framework of Order Picker Productivity, Shift Scheduling, and Staging Lane Usage.* Rotterdam: Erasmus University.

Dean, S. (2016). Multiple Regression Prediction & Confidence Intervals Demystified. *ICEAA Professional Development & Training Workshop.*

Dettling, M. (2020, 03 12). *Applied Time Series Analysis.* Retrieved from https://ethz.ch/content/dam/ethz/special-interest/math/statistics/sfs/Education/Advanced%20Studies%20in%20Applied%20Statistics/course-material/time-series/ATSA_Script_v160527.pdf

Ernst, A., Jiang, H., Krishnamoorthy, M., & Sier, D. (2004). Staff Scheduling and Rostering: A Review of Applications, Methods and Models. *European Journal of Operational Research, 153*(1), 3-27. doi:https://doi.org/10.1016/S0377-2217(03)00095-X

Evans, J. (1996). *Straightforward Statistics for the Behavioral Sciences.* Pacific Grove, CA: Brooks/Cole Publishing.

Faraway, J. (2010). Generalized Linear Models. In P. Peterson, E. Baker, & B. McGaw, *International Encyclopedia of Education 3rd Edition* (pp. 178-183). doi:https://doi.org/10.1016/B978-0-08-044894-7.01331-2

Gianazza, D. (2017). Learning Air Traffic Controller Workload from Past Sector Operations. *TM Seminar, 12th USA/Europe Air Traffic Management R&D Seminar*. Retrieved from https://hal-enac.archives-ouvertes.fr/hal-01592233/

Gomes, C., Almada-Lobo, B., Borgers, J., & Soares, C. (2012). Integrating Data Mining and Optimization Techniques on Surgery Scheduling. *International Conference on Advanced Data Mining and Applications, 7713*, 589-602. doi:https://doi.org/10.1007/978-3-642-35527-1_49

Hur, Y., Bard, J., & Frey, M. (2019). An investigation of shift and break flexibility with real-time break assignments using a rolling horizon approach. *Flexible Services and Manufacturing Journal, 31*(1), 174-211. doi:https://doi.org/10.1007/s10696-018-9305-2

Hyndman, R. (2018). *Forecasting: Principles and Practice.* OTexts. Retrieved from https://otexts.com/fpp2/index.html

Kortbeek, N., Braaksma, A., Burger, C., Bakker, P., & Boucherie, R. (2015). Flexible Nurse Staffing Based on Hourly Bed Census Predictions. *International Journal of Production Economics, 161*, 167-180. doi:https://doi.org/10.1016/j.ijpe.2014.12.007

Liao, S., van Delft, C., & Vial, J.-P. (2013). Distributionally robust workforce scheduling in call centres with uncertain arrival rates. *Optimization Methods & Software, 28*(3), 501-522. doi:http://dx.doi.org/10.1080/10556788.2012.694166

Meijsen, J. (2019, Februari 08). *Marktaandelen: Lidl wint, AH verliest*. Retrieved 09 04, 2019, from Distrifood: https://www.distrifood.nl/branche-bedrijf/nieuws/2019/01/marktaandelen-lidl-wint-ah-verliest-101121483

Meints, P. (2017, Februari 13). AH Slides Hackaton UT.

Mierswa, I. (2020, 04 18). *Why doesn't Auto-Model use Parameter Optimization for Deep Learning models?* Retrieved from RapidMiner Community: https://community.rapidminer.com/discussion/56391/why-doesnt-auto-model-use-parameter-optimization-for-deep-learning-models

Millán-Ruiz, D., Pacheco, J., Hidalgo, J., & Vélez, J. (2010). Forecasting in a Multi-Skill Call Centre. *International Conference on Artificial Intelligence and Soft Computing. ICAISC, 6114*, 582-589. doi:https://doi-org.ezproxy2.utwente.nl/10.1007/978-3-642-13232-2_71

Montgomery, D., & Runger, G. (2011). *Applied Statistics and Probability for Engineers Fifth Edition.* John Wiley & Sons Pte Ltd.

Mul, J., Poler, R., García-Sabater, J., & Lario, F. (2006). Models for Production Planning Under Uncertainty: A Review. *International Journal of Production Economics, 103*, 271-285. doi:10.1016/j.ijpe.2005.09.001

Musliu, N., Schaerf, A., & Slany, W. (2004). Local search for shift design. *European Journal of Operational Research, 153*(1), 51-64. doi:https://doi.org/10.1016/S0377-2217(03)00098-5

Nilssen, E., Stølevik, M., Johnsen, E., & Nordlander, T. (2011). Multi-Skill Shift Design for Norwegian Hospitals. *Journal of Applied Operational Research, 3*(3), 137-147.

Pakpoom, P., & Charnesethikul, P. (2018). Cyclic Personnel Scheduling with Uncertain Demand and Double Shift Requirements. *AIP Conference Proceedings, 1982*(1), 020035-1 - 020035-8. doi:https://doi.org/10.1063/1.5045441

Pearce, T., Zaki, M., Brintrup, A., & Neely, A. (2018). High-Quality Prediction Intervals for Deep Learning: A Distribution-Free, Ensembled Approach. *Proceedings of the 35th*

International Conference on Machine Learning. doi:https://arxiv.org/abs/1802.07167v3

PennState Eberly College of Science. (2020, 03 12). *Applied Regression Analysis*. Retrieved from Autocorrelation and Time Series Methods: https://online.stat.psu.edu/stat462/node/188/

RapidMiner. (2020, 04 20). *Deep Learning*. Retrieved from RapidMiner Documentation: https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/neural_nets/deep_learning.html

RapidMiner. (2020, 04 20). *Generalized Linear Model*. Retrieved from RapidMiner Documentation: https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/functions/generalized_linear_model.html

RapidMiner. (2020, 04 20). *Gradient Boosted Trees*. Retrieved from RapidMiner Documentation: https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/trees/gradient_boosted_trees.html

RapidMiner. (2020, 04 20). *Performance (Regression)*. Retrieved from RapidMiner Documentation: https://docs.rapidminer.com/latest/studio/operators/validation/performance/predictive/performance_regression.html

Rijksoverheid. (2019, 11 08). *Welke contracten zijn er voor oproepkrachten?* Retrieved from Rijksoverheid: https://www.rijksoverheid.nl/onderwerpen/arbeidsovereenkomst-en-cao/vraag-en-antwoord/welke-contracten-zijn-er-voor-oproepkrachten

Rijksoverheid. (2019, 11 08). *Wet arbeidsmarkt in balans (WAB): wat er verandert vanaf 1 januari 2020*. Retrieved from Rijksoverheid: https://www.rijksoverheid.nl/onderwerpen/arbeidsovereenkomst-en-cao/plannen-kabinet-voor-meer-balans-tussen-vast-werk-en-flexwerk

Ruiz-Aguilar, J., Moscoso-López, J., Turias, I., & González-Enrique, J. (2017). Forecasting Freight Inspection Volume Using Bayesian Regularization Artificial Neural Networks: An Aggregation-Disaggregation Procedure. *Intelligent Systems and Computing, 649*, 179-187. doi:https://doi-org.ezproxy2.utwente.nl/10.1007/978-3-319-67180-2_17

Seada, A., & Eltawil, A. (2015). Modeling and Analysis of Workforce Management Decisions in Modern Call Centers. *International Conference on Industrial Engineering and Operations Management* (pp. 1-10). Dubai, United Arab Emirates (UAE): Institue of Electrical and Electronics Engineers. doi:10.1109/IEOM.2015.7093761

Serengil, S., & Ozpinar, A. (2017). Workforce Optimization for Bank Operation Centers: A Machine Learning Approach. *International Journal of Interactive Multimedia and Artificial Intelligence, 4*(6), 81-87. doi:10.9781/ijimai.2017.07.002

Simeunović, N., Kamenko, I., Bugarski, V., Jovanović, M., & Lalić, B. (2017). Improving Workforce Scheduling Using Artificial Neural Networks Model. *Advances in Production Engineering & Management, 12*(4), 337-352. doi:https://doi.org/10.14743/apem2017.4.262

Tan, P., Steinbach, M., & Kumar, V. (2013). *Introduction to Data Mining (New International Editon).* Pearson Education Limited. doi:10.1017/CBO9781107415324.004

van Hulst, D., den Hertog, D., & Nuijten, W. (2017). Robust Shift Generation in Workforce Planning. *Comput Manag Sci, 14*, 115-134. doi:https://doi-org.ezproxy2.utwente.nl/10.1007/s10287-016-0265-2

9. Appendix

All Appendixes are removed since they contained confidential information.

A. Context Analysis Interviews

B. Capacity Planning Brainstorm at DCO (11th December 2019)

C. Data Preparation: Colli Predictions

D. Data Preparation: Realized Order Picker Productivity, Picking Hours, and Picked Colli

E. Data Preparation: Absence Due to Illness

F. Hypothesis Testing

G. Modelling: Directly Predicting the Required Number of Flex Employees

H. Modelling: Predicting #Colli, then #Flex

I. RapidMiner Process: Generating Prediction Intervals

J. Opening RapidMiner Files on Own Computer