Vaasan yliopisto
UNIVERSITY OF VAASA

Luukas Kuusrainen

Suitable demand forecasting method for stock quantity optimization in the food industry

School of Technology and Innovations
Master's thesis in Technology
Industrial Systems Analytics

Vaasa 2023

UNIVERSITY OF VAASA
School of Technology and Innovations
Author: Luukas Kuusrainen
Title of the Thesis: Suitable demand forecasting method for stock quantity optimization in the food industry
Degree: Master of Science
Programme: Industrial Systems Analytics
Supervisor: Petri Helo
Year: 2023 **Number of pages:** 77

ABSTRACT:
The goal of this master's thesis is to answer the research question, which is “what is the suitable way for the food industry to make demand forecasting?”. In this research, various demand forecasting methods were compared using a dataset from the food industry. These methods included Facebook Prophet, Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and 90 days average, which is the research baseline. Forecasting methods are compared using Root Mean square error (RMSE) calculation. The results showed that the 90-day average method performed the best in accuracy. However, it is essential to note that other factors, such as data collection and unusual demand changes, can impact the effectiveness of demand forecasting in the food industry. The study also discussed the importance of inventory management in the food industry and the impact of stock quantity optimization, including using demand forecasts to optimize stock quantities. Overall, this research provides insights into demand forecasting in the food industry.

KEYWORDS: Demand forecasting, Food industry, Stock quantity optimization

Contents

Figures

Tables

Algorithms

Abbreviations

MAE	- Mean absolute error
MSE	- Mean square error
RMSE	- Root mean square error
MAPE	- Mean absolute percentage error
CNN	- Convolutional neural network

LSTM	- Long short-term memory

1 Introduction

This section presents the reason for the research. The background shows the current situation and ways of doing things. The research question presents that there is room for improvement and the limitations show the research environment. The research plan tells how the problem will be solved and how it can be approved.

1.1 Background

Demand knowledge for all industries is essential. It can be used to optimise the stock quantities and for many other parts of manufacturing as production planning, budgeting, working schedules for employees etc. (Folorunso, Adewale , Adewale , & Okesola , 2011)

In the need for estimated demand, there can be significant differences in the length and accuracy of the demand. These demands are often made by hand or using an average from the sales history. There are also more sophisticated methods for making demand forecasting, such as moving average calculation, Neural network-based solutions, and other algorithms. (Berry & Haile, 2021)

In the food industry, expiration is an issue. Often, the time for expiration is short. The food industry also has rapid demand changes, making demand forecasting difficult. The amount of products in the food industry is also high, so it is very work-intensive to make good forecasts by hand. Correctly used accurate demand forecasts would lower waste and increase the food industry's profitability. (N. de P. Barbosa, 2015)

Demand forecasting benefits the food industry in multiple different ways. Optimised stock levels provide better customer availability and fresher products with better expiration dates. Lower inventory levels are easier to handle and give suppliers and grocery stores more space. This provides the physical space for a broader selection of products and means lower capital assets tight to the inventory, which can be invested in the broader selection. (Adebanjo & Mann, 2000)

Comparing data-driven decision-making for controlling the stock quantities is a major possibility in this case. The importance of this possibility contains many benefits. Benefits can be improved security of supply, a smaller amount of waste because of obsolescence, fewer capital assets tight to the warehouse, less used space for stocks in the warehouse and fresher products delivered for the end customer. This research presents the needed base work towards data-driven decision-making to control stock quantities. This is what makes this research important for the companies in the food industry that still need to implement data-driven stock quantity optimisation methods.

1.2 Research question

The research question is "what is the suitable way for the food industry to make demand forecasting?". The answer to this question cannot be short because there are many different ways to define how accurate the demand forecast is. This research should also present how the most suitable demand forecasting method is selected for the food industry.

1.3 Research delimitations

This research is limited to selecting the demand forecasting method and does not include the data collection for the historical sales data. How the demand forecast is used is also critical to decreasing waste and increasing profitability. However, this research does not try to find suitable solutions for using demand forecast results to decrease waste.

The need for food industry stock quantity optimisation is for a short period of time, so the forecast accuracy calculation is limited to 30 days.

There are many different age products in the food industry, and there are different issues with products with a short period of sale data. This research is limited to products that are older than a year.

Demands between different products vary a lot, and in smaller grocery stores, products might not have sales for extended time periods, or the sales might be a couple of pieces a day. On the other end, some food factories might sell tens of thousands of products each day. This research limits forecasting products sold at least over ten pieces every day.

In the food industry, there are many different size companies and operators. This research does not limit to any specific size. This research only includes food factories and grocery stores.

Food production and grocery stores contain many types of stocks and demand, so this research limits researching the suitable demand forecasting method for the ready products.

The demand forecasting accuracy should be measured for each day because the food industry does not need information on how the demand changes in one day. However, it should be one or a couple of days from a stock quantity perspective. Demand changes a lot between the weekdays and weekend days. In the grocery store, the demand can be much higher on a weekend day than on a weekday, but there might not be production at all in a food factory on weekends and ether there are sales. That is why forecast accuracy on a weekly average is not enough, and hourly is unnecessary. For these reasons, the research limits demand forecasting to daily sums.

This research is limited to stable products without known significant effects on sales during the test period. These known changes can be campaigns, discounts, significant price changes or product placement changes. Products in the test set should not include these kinds of changes.

The research tries to define a suitable demand forecasting method for the food industry in Finland. The export sales can be considered for the food factories, but the grocery store data and environment perspective are focused on Finland.

This research aims to find the most suitable demand forecasting method for the food industry but this research studies the accuracy of demand forecasts with only the basic functionalities. Earlier research shows that there are many other features which could be important in the future of the demand forecasting method but not yet in this research. That is why the requirements table categorises the requirements as critical for this research and essential for future proving. Inside these categories are three different priority sections which present the importance with low, medium, and high for this research and future proving priority.

Table 1. Requirements table for this research and features which might be important for future approving.

	Priority in this research			Priority for future proving		
Feature	low	medium	high	low	medium	high
Yearly seasonality			x			x
Weekly seasonality			x			x
Multiple input variables	x				x	
Holidays	x					x
Campaign data	x					x
Weather data	x					x
Daily forecast		x				x
Weekly forecast			x			x

This requirements table presents well how many features are essential for future approving but the forecasting process but this research studies the basic functionalities of demand forecasting. This is why many of these features are not yet needed in this study because they can be easily excluded from the dataset used in this study. However, they should be included in theoretical research to give a better understanding to someone using the results of this research.

1.4 Research plan and structure

Research starts by defining the needs and environment for the food industry demand forecasting. The research aims to make forecasts for stock quantity optimisation, so the feasibility of that is also discussed in the first section. The second section discusses the different theories of inventory management and demand forecasting theories. These vary from different inventory models to understanding the demand and background of demand forecasting. One major part of this section is the comparison methods of forecasts.

The third research section defines multiple demand forecasting methods for comparing and studying these forecasting methods and their feasibility for the food industry. It also discusses details about the difficulties of demand forecasting in the food industry.

The results section is the fourth section, and this section presents how the different methods are used in practice and compare the results from different methods against each other. We can see from the following picture how the comparison process goes.

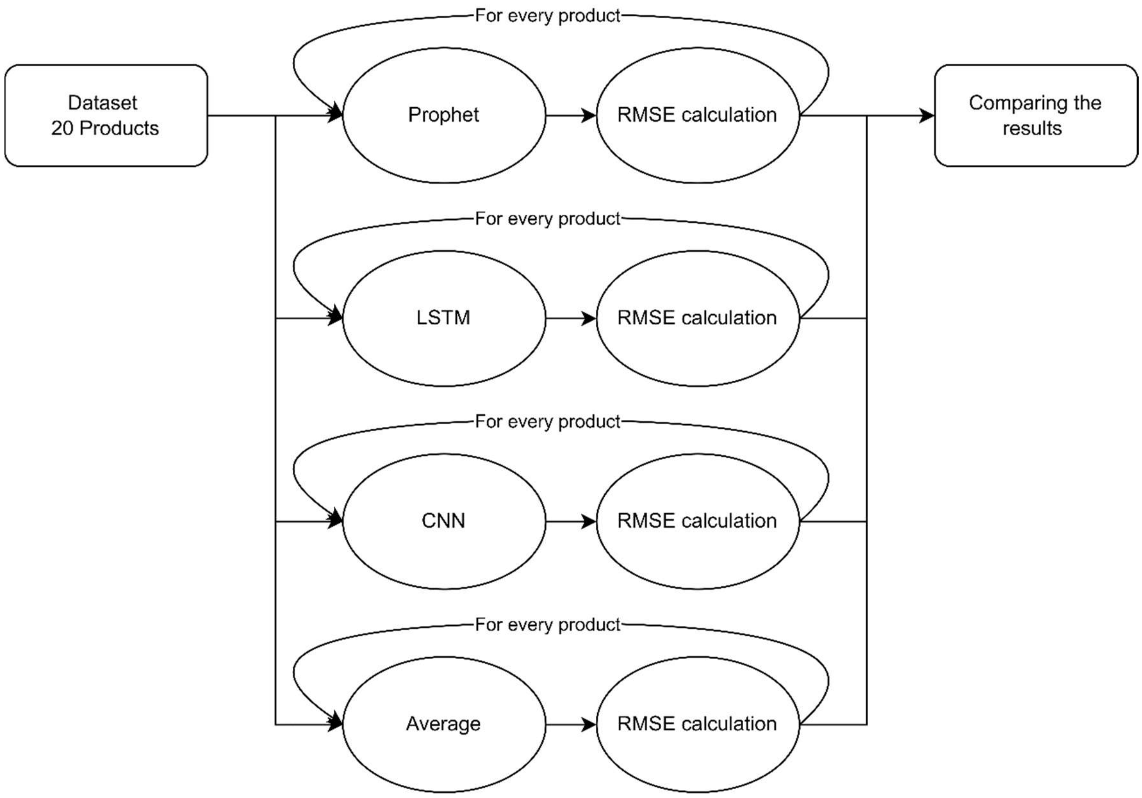

Figure 1. Research process architecture with presents 80 different forecasts.

From the figure above can be seen that every forecast method makes a forecast for every product which there are 20. Four different methods make a total of 80 forecasts. The results section also presents the error calculation from every method.

The conclusion is the last section which combines the results and discusses the future results opportunities.

2 Demand and inventory management in food industry

The primary use for demand forecasting is optimising stock quantities. To understand the purpose of demand forecasts and how inventory models consider the demand forecast when ordering products into stock.

2.1 Inventory models

There are multiple variables which inventory models relate to. Understanding the meaning of different variables is essential to understand the inventory models. Demand is an important variable of inventory models, and understanding what the demand will be for the time period the inventory is planned to last. Delivery lead time is another critical variable. If the lead time changes, this needs to be considered because it changes the time period length of what the inventory is planned to last. When the delivery time is constant is easy to take consider when placing the order. (Fildes, Ma, & Kolassa, 2019)

In some cases, the supplier might provide regular delivery dates and times when the order needs to be placed so that the order is included in the delivery. This is often the case in the food industry and is easy to consider. Inventory models should take into consideration the costs associated with inventory. Inventory holding costs consist of many different costs as the capital costs tight up to the inventory, and expiring products cause waste and handling costs. There can be ordering costs for placing the order or delivering costs. These variables should be taken into consideration when choosing the inventory model. (Fildes, Ma, & Kolassa, 2019)

2.1.1 Continuous ordering model

When using a continuous order model, the order sizes are the same. The order cycle length changes, as we can see in figure 1.

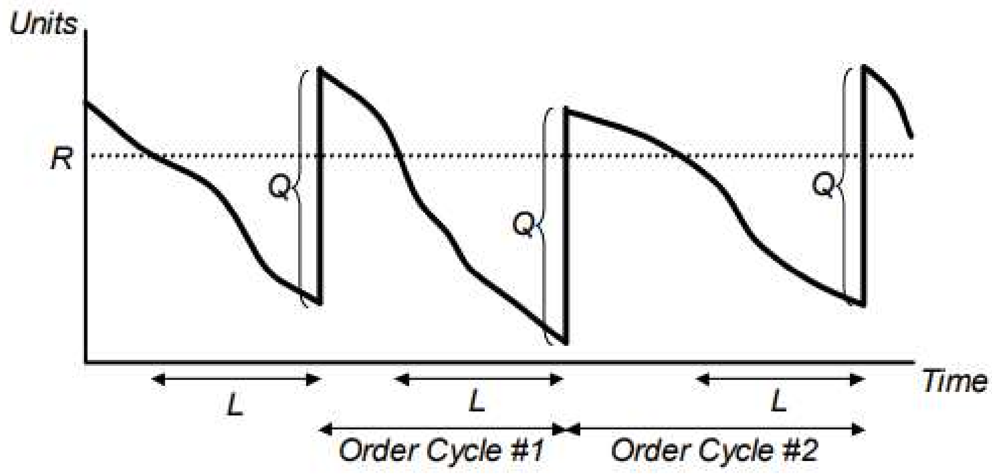

Figure 2 Continuous inventory model (Willems, 2001).

When the inventory level goes under the ordering limit (R level in figure 1), the same amount (Q in figure 1) of products is ordered. The ordering limit should consider the delivery length so that the product does not run out before the delivery arrives. This model works well when the supplier defines the order size and the ordering party hasty order that exact quantity. In the food industry, the production batch sizes can also be exact, so this model can also be used for the food manufacturing industry. Ordering small mounts makes sense when the product's value is high, and demand is low. (Willems, 2001)

2.1.2 Periodic ordering model

Reordering the product periodically means that the order cycle length is always the same, but the ordering size changes with the demand. An example of this we can see in figure 2.

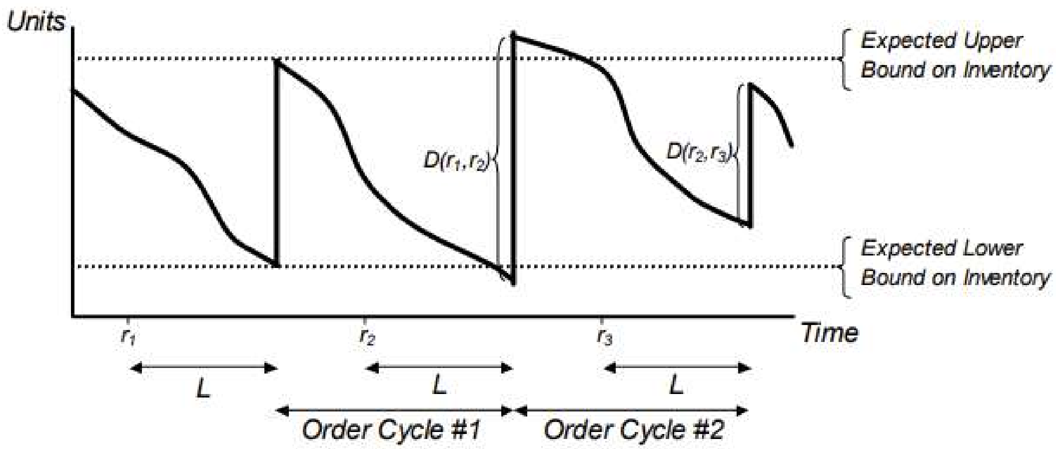

Figure 3 Periodical inventory model. (Willems, 2001).

When making periodic orders, there should be an expectation for stock quantity at the delivery time, and an expedition on how much sales will happen during the next ordering cycle. The order should be the sales in the next ordering cycle, and if it is assumed that on the delivery time, the safety stock would be too low or too high, it should be taken into account when making the order. This ordering model suits well for products whose supplier has fixed delivery dates. This is often the case in the food industry because the products are delivered to multiple customers at once, and there are predefined logistics routes. (Willems, 2001)

2.2 Demand in food industry

2.2.1 Seasonality

Products in the food industry have strong seasonality. Seasonality patterns are predictable. Pattern lengths vary from lengths of weeks to yearly seasonality. Demand changes during the week can be significant. Usually, demand on weekdays is much lower than on weekend days Same happens during the year, and demand is much higher during the summer than in winter. (Fildes, Ma, & Kolassa, 2019)

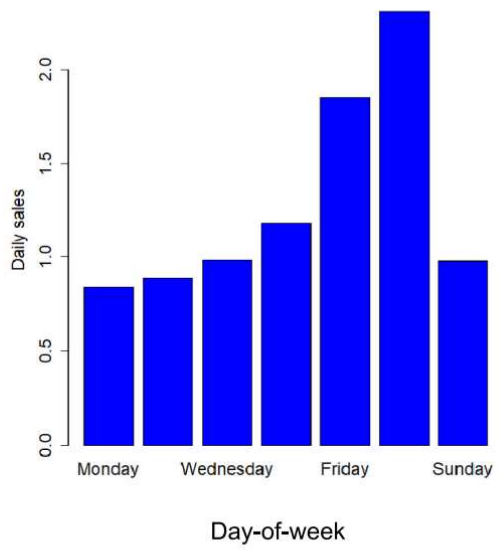

Figure 4. Weekday distribution of beer sales presents well the difference between Sunday to Thursday and Friday to Saturday. (Fildes, Ma, & Kolassa, 2019).

The importance of taking into account the weekly distribution of sales is easily seen in the earlier study, which is presented in Figure 3. The difference Between the highest weekday and the lowest weekday is more than two times the sales.

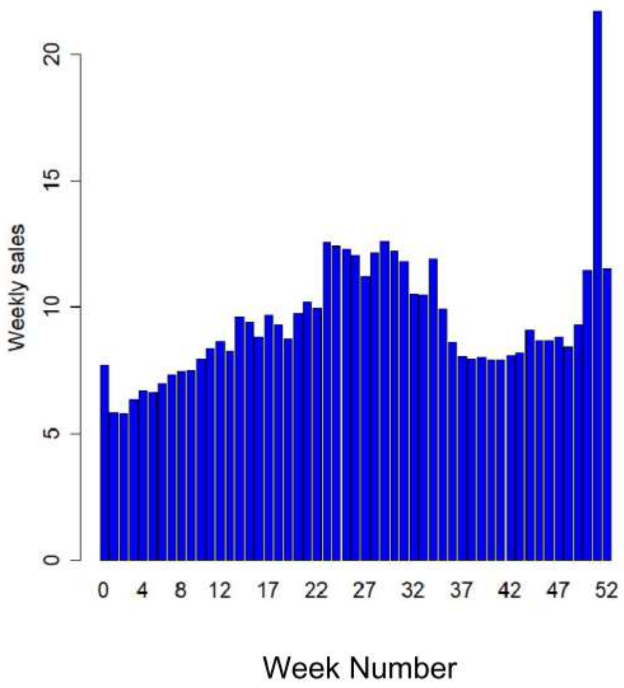

Figure 5. Sales aggregated to weeks sales present well the difference between different seasons. (Fildes, Ma, & Kolassa, 2019).

Weekly aggregated sales in Figure 4 present the yearly seasonality well, and we can see the difference between winter and summer weekly sales. The difference between the lowest and highest weeks is over double, even when leaving the special winter events out.

2.2.2 Special events

Special events have a significant effect on the food industry. Special events depend on a lot about the country. In Finland, the calendar events are, for example, Christmas, mid-summer fest and Easter. Comparing this to earlier research in Figure 4, which has been done in the United Kingdom, we can easily see the effect of Christmas. The sales over double during a short period and come back to an even lower level after the special event. If this is not taken into account in the food industry, it will cause a significant loss in sales because the stocks would be empty. Drop after special events can also cause significant waste, and this needs to be also considered. (Fildes, Ma, & Kolassa, 2019)

Special events can also be local events such as festivals, sports events or other significant events in local areas that affect customer behaviour. Looking into national sales studies does not show these events, but for grocery stores, these can have a significant effect on sales, which is why this needs to be taken into account at the local level. (Fildes, Ma, & Kolassa, 2019)

Special events can also be holidays. There are multiple holiday seasons during the year in Finland, and it is seen in the study Fieldes, Ma, and Kolassa have done that the demand rises during holidays. The difficulty in considering the holidays in demand is that the holidays might move to different days between different years. Therefore it is essential to understand the effect of the holiday and its timing to correct the time period.

Religious events also affect sales, but this is a minor effect in Finland. (Fildes, Ma, & Kolassa, 2019)

2.2.3 Weather

The effect of weather on sales depends a lot on the product. What is discovered in earlier research is that, for example, drink sales have a significant relationship to weather. Higher temperature does not mean every time higher sales, and some products can sell sale more at lower temperatures as hot drinks do. The temperature effect is not linear to sales. When the temperature goes from average to high, it increases the sales of soft drinks, but in other ways, the temperature does not affect soft drink sales. This makes it much more challenging to take into account the temperature. (Fildes, Ma, & Kolassa, 2019)

Weather is not only about the temperature. Other weather changes than temperature can be rain, cloudiness, snowing and sunshine. Different changes and combinations of water have different effects on demand. Because of the non-linear effect on sales by the temperature and other effects of the weather on sales, there should be a definition for product or product group that what kind of weather combination has which kind of effect on sales. Fildes, Ma and Kolassa find out in their research that this information would improve the understanding of demand forecasting by 10 per cent. (Fildes, Ma, & Kolassa, 2019)

Weather forecasts are easily available, but accuracy in more extended time periods needs to be more accurate to be used in demand forecasting. This is why weather should be used in the food industry to make more accurate short-term demand forecasts, but for long-term demand forecasting, it is not necessary. (Fildes, Ma, & Kolassa, 2019)

2.3 Stock quantity optimisation by demand forecasts

Industries contain many kinds of stocks. There can be storing raw materials ready for production, ready products to be delivered, products on store shelves and many more. All these stock quantities can be optimised, but this research aims to optimise ready products from factories to end customers in the food industry. (Priymasiwi, 2018)

The stock quantity is controlled by ordering raw materials, products, or production. Sales consume the stock quantity. In the food industry, waste can also consume the products from the stock. If expire day passes, then the product is wasted from the stock. These are the three main effects on stock quantities in the food industry. The optimal stock quantity is low as possible without missing delivering any sales. The stock quantity cannot be optimal because of variance in these three effects. (Priymasiwi, 2018)

2.3.1 Adding quantities to stock

Factories add products to the ready product stock by controlling production. Stores add products to their stock by ordering more products from the supplier. Both of these ways have a lead time. There can be significant differences in lead times between different products. Lead time variance and uncertainty are significant issues for stock quantity optimisation. The employee schedules and raw material stock affect the lead time in a food factory. (King, 2011)

There can also be variances in the production or order size. The produced or delivered quantity might be different from the ordered quantity. In the worst-case scenario, there might not be delivery or production due to production line issues, material shortages, or logistics problems. In these cases, the safety stock should cover the usual happening variations in deliveries and production. (King, 2011)

2.3.2 Consuming the stock quantities

Consuming stock quantities happens by selling the products or in material stocks by production consuming the material. The time period between the order and the stock consumption is usually short in the food industry, which is why demand forecasting would help provide information about the demand. If the stock does not contain enough quantity, the sales have been lost, affecting profitability. (A. Olsson, 2008)

In a grocery store, the stock is consumed when a customer purchases something. The sale is lost if there is no product in the store and the shelf is not empty. If too many products are on the shelf, the grocery store might sell the products with a discount so that they would not expire, but this also affects the profitability of the sale. (A. Olsson, 2008)

Another form of consuming stock quantities is waste. This is a significant issue for grocery stores and food factories in the food industry. Usually, what causes the waste is expiring products, but wrong storing temperatures might cause products to be wasted. (Alexander, et al., 2017)

2.3.3 Waste and security of supply optimisation

The demand forecast is used for waste and security of supply optimisation in the food industry. An accurate demand forecast is important because, in the food industry, the products usually have an expiring date. Changes in demand are sudden but also partly presumable. That is why the average would be too late and cause significant waste when demand drops significantly. (Silva, Figueiredo, & Braga, 2019)

The food industry in Finland is the largest food waste producer. Between 10-15 % of the food produced in Finland goes to waste, which means 385 -485 million kilograms.

Finnish food factories cause 20 %, and grocery stores cause 18 % of the waste. The rest of the waste happens in early production, in restaurants or by the consumers. (Kiuru & Silvennoinen, 2016)

There are multiple reasons why food goes to waste, but one is that products expire in grocery stores or food factories because of poor stock quantity optimisation. This research focuses on grocery stores and food factories, which cause about 38% of Finnish food waste, between 140 and 180 million kilograms. Waste causes losses, but also empty stocks cause losses in the sale. During an extended time period, these losses in sales and waste can build up to significant losses in profits. (Kiuru & Silvennoinen, 2016)

2.3.4 Safety stock calculation

Demand forecasts can be used to order products from suppliers or production in the factory. Using traditional safety stock calculation calculates the order size from the demand forecast. Safety stock calculation can calculate the variance in lead time and demand. (Lisan, 2018)

The lead time is constant in the food industry because the logistic routes are often constant, and the products are delivered constantly. The delivery time is also constant in food production because there is only a slight variance between planning and producing the product. (Lisan, 2018)

That leaves the demand deviation. Demand deviation would be unnecessary if the demand forecast were accurate. However, it is known that demand forecasting is not absolutely accurate, so that is why demand devotion should be added to the order.
The stock quantity calculation presents one major issue for the food industry because it increases stock quantities. If the safety stock is too high, it can cause products to expire and decrease profits. (King, 2011)

2.4 Demand forecasting

There is much research about demand forecasting, and this section investigates earlier research about demand forecast. Earlier research shows that there is no singular method that is more accurate in demand forecasting than all other methods. However, it also shows that combining the methods by making hybrid forecasting models presents the most promising results. Also, additional input data used as multivariable input put is making the forecasting more accurate. For example, weather which is one of the most important.

2.4.1 ARIMA and ARIMAX in demand forecasting

Earlier research shows that ARIMA and ARIMAX model has been used successfully in demand forecasting. Studies show that ARIMA and ARIMAX have successfully made short-term forecasts. The forecast length in a study that Huber, Gossmann and Stuckenschmidt have done was three days. In this study, they used fresh buns and bread. For this use case, the short-term forecast fits well, showing that ARIMA and ARIMAX contain the potential for this use case. (Huber, Gossmann, & Stuckenschmidt, 2017)

This study shows well how it is possible to use a method which does not contain a weekly seasonality component to forecasts something with high weekly seasonality. They used data only from weekdays to forecast weekdays with ARIMA. With ARIMAX, which creates a possibility to use multiple variables, they did use a variable for weekday and weekend days. This takes well into consideration the high weekly seasonality. (Huber, Gossmann, & Stuckenschmidt, 2017)

The study also presents the most accurate results they did get by combining ARIMA and ARMIAX. This combination is most accurate in their research at forecasting accuracy. (Huber, Gossmann, & Stuckenschmidt, 2017)

2.4.2 SVM demand forecasting

Research by Du, Leung Zhang, and Lai studies SVM (Support vector machine) used in demand forecasting for farm products that can expire quickly. They are making 30-day forecasts which fit the well to food industry's needs. They compare the SVM to RBF, which is a radial basis function. They find that SVM is more accurate for forecasting the demand of the RBF. (Du, Leung, Zhang, & Lai, 2011)

The research presents significant usage of multiple data sources to input and train their model with different variables, as seen in the following figure.

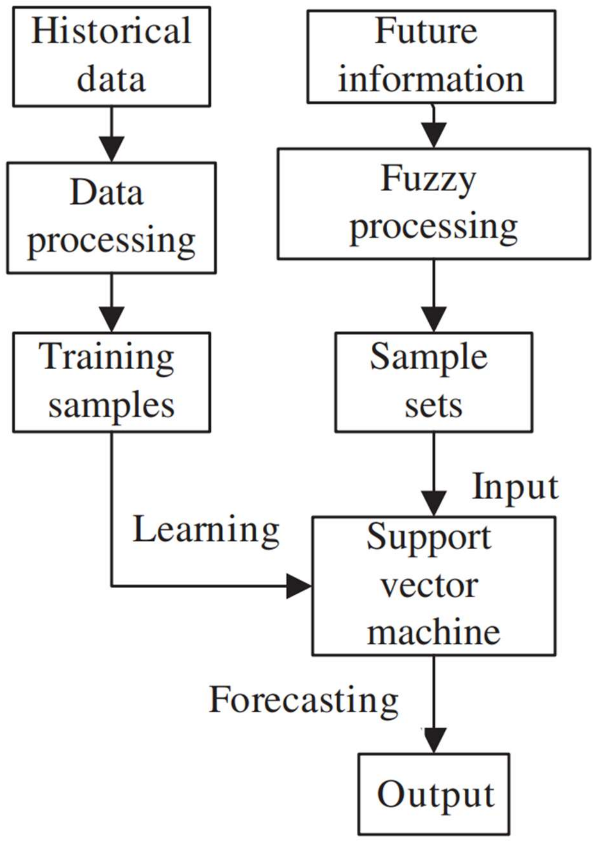

Figure 6. SVM structure using future data for more accurate forecasting (Du, Leung, Zhang, & Lai, 2011).

They have used 11 different variables as input variables for their model which are:

- Sales quantity of the day before the forecasting day.
- Sales quantity of the day 2 days before the forecasting day.
- Sales quantity of the same day in the preceding week.
- Type of the forecasting day(weekend or week day).
- Type of the day before the forecasting day (weekend or week day).
- The highest temperature of the forecasting day.
- The lowest temperature of the forecasting day.
- The highest temperature of the day before forecasting day.
- The lowest temperature of the day before forecasting day.

- Weather condition of the forecasting day. (sunny, cloudy, overcast or rainy)
- Weather condition of the day before forecasting day. (sunny, cloudy, overcast or rainy)

his listing from their research presents well how they have found 11 influential variables to support their forecast by only using two different data sources. They have used sales and weather forecasts to aggregate these 11 values. This way, they had made a 98,62 per cent accurate forecast with SVM when RBF was 97,47 per cent accurate for the same case. (Du, Leung, Zhang, & Lai, 2011)

Their study presents how data aggregation can be used for demand forecasting. Weather affects demand forecasting, and this research shows a great example of how to do it.

Previous research has also studied comparing SVM to other forecasting methods such as ARIMA RNN (Recurrent Neural Networks) and LSTM (Long Short-Term Memory). Wang and Liu have studied demand forecasting in the retail industry with data from 200 hundred products in multiple stores. They categorised the products into two categories: expiring products and non-expiring products. From these two, the expiring products are similar for the food industry. (Wang, Liu, & Liu, 2019)

Research uses five different methods to evaluate the forecasting methods:

- Runtime
- Generalization Ability
- Convenience
- Predictive Accuracy
- Cost

These variables present well the different methods which can be used to evaluate the forecasting model, which from the predictive accuracy and cost, are most interesting for the food industry and the objectives of this research. (Wang, Liu, & Liu, 2019)

The results between different methods and criteria can be seen in the following table.

Table 2. Table from ARIMA, SVM, RNN and LSTM study results (Wang, Liu, & Liu, 2019)

Perishable(normalization)	ARIMA	SVM	RNN	LSTM
Runtime	0.2854	0.5730	0.7153	0.5007
Generalization Ability	0.3333	0.5167	0.5667	0.6667
Convenience	0.3636	0.4545	0.5455	0.6364
Predictive Accurancy	87.13%	99.10%	98.90%	96.80%
Cost	0.1818	0.7273	0.6364	0.5455

The results show that the SVM is the most accurate in retail demand forecasting research. RNN is the second most accurate. The difference between SVM and RNN is only 0,2 per cent in predicted accuracy. The difference between the most accurate SVM and the third accurate LSTM is 2,3 per cent. The worst predicted accuracy is with ARIMA. Between SVN and ARIMA, the difference is 11,97 per cent. This research presents how a difference in predicted accuracy affects cost savings because even the difference between the most accurate and second accurate is only 0,2. However, it is a 14 % difference in cost savings. (Wang, Liu, & Liu, 2019)

2.4.3 Regression model used to demand forecast deviation

A study done by Donsealaar, Peters, Jong and Broekmeulen presents how different promotions can be considered in forecasts. This study uses a regression model to forecast different campaigns. In their research, they researched many different products from the food industry and did find out that products sell on average 14 times more on campaign week than on a typical week. They used a regression model to forecast this higher sale during the campaign. (Van Donselaar, Peters, De Jong, & Broekmeulen, 2012)

They applied three different regression models categorised, linear and quadratic. The categorised model used different discount percentage caps 0-10,10-20…50-60 per cent discounts. For linear and quadratic models, they used the discount percentage as it is. The best-performing model was a quadratic model. They did not find out that the categorisation of percentages with dummy variables would improve the accuracy of the demand forecast. (Van Donselaar, Peters, De Jong, & Broekmeulen, 2012)

In their study, they also did find out that they got different results with a different types of products. Stable products had better results and should be analysed separately from non-stable products. Non-stable products could benefit also including products from other product categories for analysis. (Van Donselaar, Peters, De Jong, & Broekmeulen, 2012)

2.4.4 Neural network in demand forecasting

There is existing research using NN (Neural networks) for demand forecasting. Research shows that there is a benefit to using NN compared to ARIMA. One of the main objectives is combining data with other features to make more accurate forecasts. Research done by Chen and Ou presents how whether can be used as input for a neural network to make more accurate forecasts. (Chen & Ou, 2008)

Their research used 11 months of sales data and a similar amount of weather data. In their study, there are similarities between the sales data and weather data which can be seen in the following pictures.

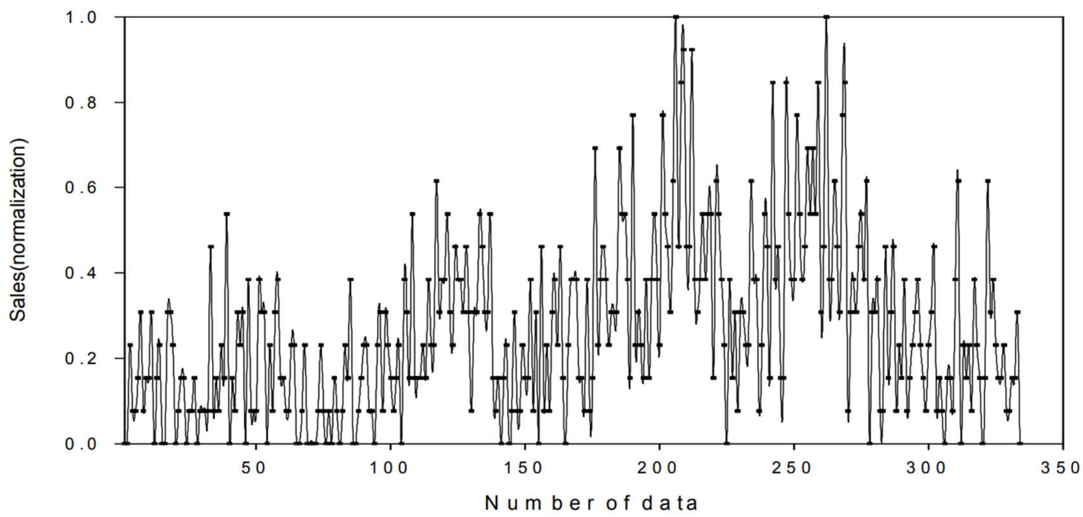

Figure 7. NN study sales quantity data (Chen & Ou, 2008)

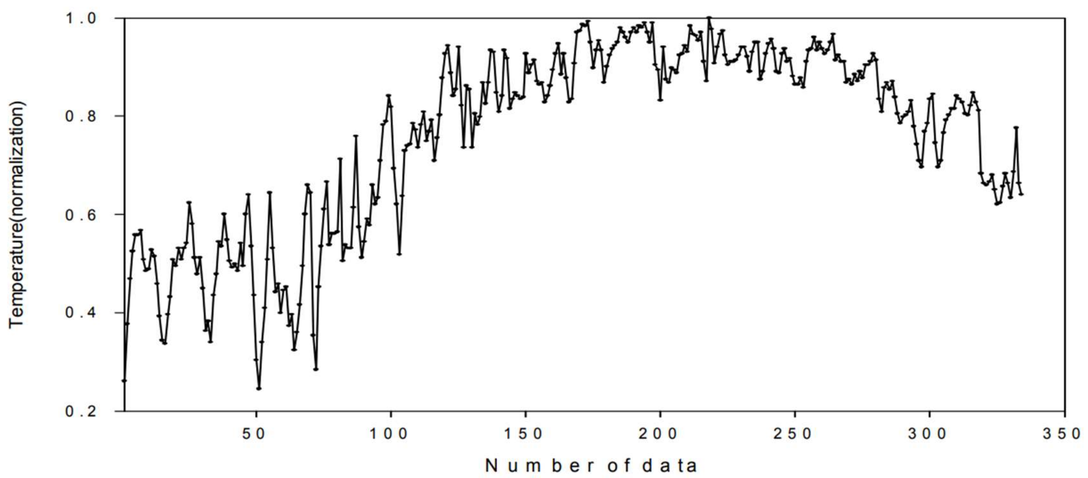

Figure 8. NN study weather temperature data (Chen & Ou, 2008)

Similar seasonality trends can be found in both sales and weather data points. Using this kind of correlating data as input data for NN, Chen and Ou were applied to get more accurate results than their baseline, which was ARIMA. They used 120 days as the test period, and the ARIMA MSE (Mean Square Error) was 0.03653, whereas the NN model MSE was 0.01977. This research shows how additional information can significantly improve demand forecast accuracy using NN. (Chen & Ou, 2008)

This research also presents the importance of parameters when using NN. Chen and Ou found that the most critical parameter for NN forecasting is the count of input nodes of the NN model. (Chen & Ou, 2008)

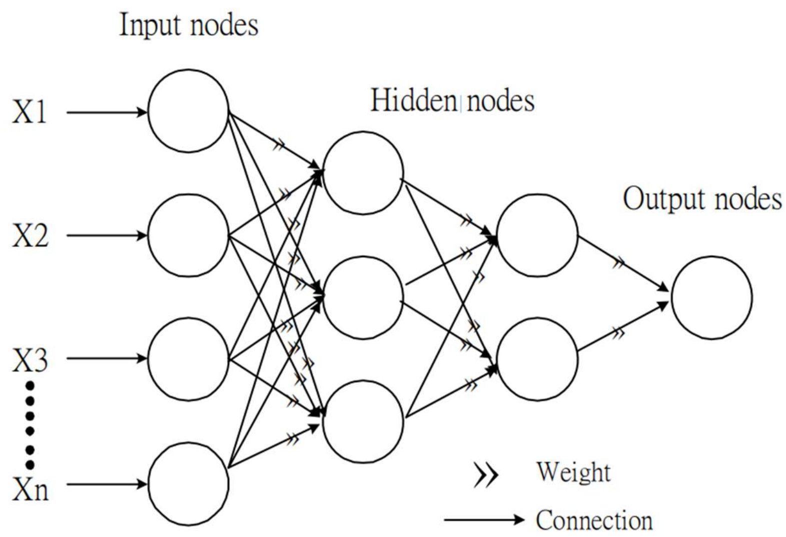

Figure 9. Neural network structure (Chen & Ou, 2008)

As shown in the neural network structure picture, the input node amount also affects the count of hidden nodes, highlighting the importance of the input node count as a parameter. Chen and Ou tested five different input node counts from 3-8. The difference between these in accuracy was significant. The eighth was the most accurate MSE 0.01977, and the worst was with three input nodes MSE 0.03697. This shows the importance of input node count as a parameter. (Chen & Ou, 2008)

2.5 Comparing forecasts

Comparing different forecasts can tell which of the forecasting ways is the optimal way to do forecasting. This contains two significant issues. The first one is that there is no possibility of knowing the future demand. That is why the forecast will be done for some time period where the demand is known. There should be multiple time periods, so that the demand forecast does not overfit to one time period. The second problem of comparing is how to calculate the error and what the most optimal forecast is. (Aijaz & Agarwal, 2019)

2.5.1 Time period selection

Error analysing time length should be the length that is meaningful for the food industry. The food industry often orders or produces products periodically. For grocery stores and food factories, the length of the time period is different and usually much shorter for the grocery stores because they can have deliveries in a short period of time, and they do not need to schedule their employees by that. However, in a food factory, the time period is extended so that they can plan their working schedules and raw materials ahead. This is why the length of the demand forecast is chosen to be 30 days so that it is long enough for grocery stores and food factories. (Khair, Fahmi, Hakim, & Rahim, 2002)

The demand changes a lot during the year, and events like holidays and seasonality change, which can significantly affect the demand. This is why the time of the year of the testing period is challenging to select. There is not only one time period selected for this research, but there are ten time periods. These ten time periods lay out evenly throughout the year and should present the best possible demand forecasting method for any time of the year. This causes the testing period to be more than a year ago. (Khair, Fahmi, Hakim, & Rahim, 2002)

2.5.2 Error calculation

The Absolut error of one forecasted value is the difference between the forecasted and known value for the same timestamp. One absolute error value for one timestamp does not tell much about the performance of the forecast. That is why error values for different times need to be calculated to one number, which can be compared between different forecasts. (Koponen, Ikäheimo, Koskela, Brester, & Niska, 2020)

There are multiple ways to calculate the error. The main difference between different error calculation methods is the effect of the error, linear or non-linear. Linear error calculation uses absolute error, and the non-linear uses squared error or other non-linear functions. (Khair, Fahmi, Hakim, & Rahim, 2002)

2.5.2.1 MAE Mean absolute error

MAE is shortened of mean absolute error. Error is calculated between every forecasted data point and known data point for the same timestamp in the selected time window. The average is calculated from error values. Error values are always absolute, so the average does not combine negative and positive error values. (Pascual, 2018)

$$MAE = \frac{1}{n}\sum |\gamma - \hat{\gamma}|$$

In this example picture, the forecasted model is the green linear line, and the existing values are the blue ones. Red vertical lines are absolute error values. (Pascual, 2018)

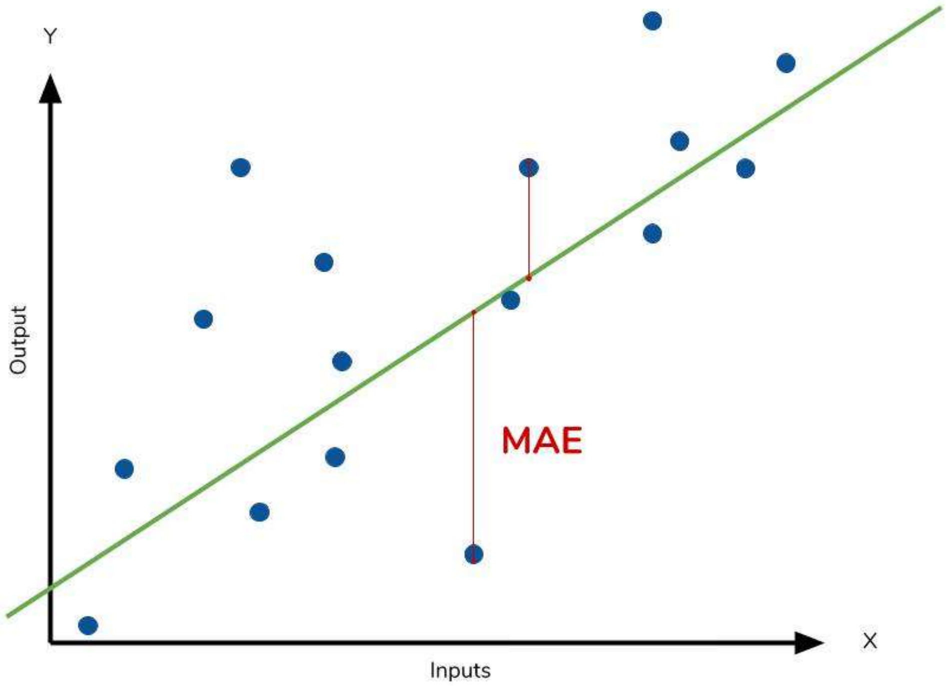

Figure 10. The picture presents the difference between the forecasted model and sales history (Pascual, 2018).

2.5.2.2 MSE Mean square error

MSE is shortened of mean square error. This is a non-linear way to calculate the error. The square in the formula causes non-linearity. That means the difference in error is not valued the same depending on how near it is from the optimal value. This is beneficial in the food industry because it appreciates the option which tries to spread the error spikes. This is a wanted effect in the food industry to avoid wastage. (Pascual, 2018)

$$MSE = \frac{1}{n}\sum (y - \hat{y})^2$$

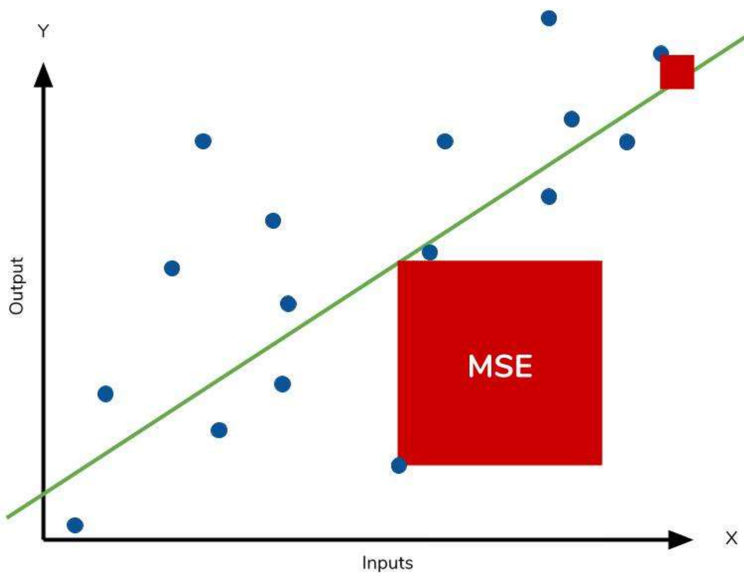

Figure 11. Red box presents the squared error between the forecasted model and known datapoint (Pascual, 2018).

2.5.2.3 RMSE Root Mean square error

RMSE is shortened for root mean square error. RMSE includes the MSE calculation, and because the individual error values are squared, the error is non-linear in MSE. The individual error values are treated the same as in MSE error calculation. The mean result is divided by the count of the measurements and then by the root. (Pascual, 2018)

The difference between MSE and RMSE is that RMSE is at the same unit as the original forecasted time-series value. RMSE gives the same result, which would be the best option than MSE. However, it is easier for humans to understand that what is the difference between two different models when the error is divided by the count of the data points and in the root so that the unit is the same because RMSE includes the benefits of the MSE and is also easy to understand in the result section of this master thesis. (Pascual, 2018)

2.5.2.4 MAPE Mean Absolut percentage error

MAPE is shortened of mean Absolut percentage error. This is a linear way to calculate the error. The benefit of the MAPE is that the output is in percentage, making it easier for the user to understand. The percentage of the error does not consider the size of the demand forecast, and that gives the possibility to compare the accuracy between the different sizes of sale products, but this is not often possible because the variance in smaller demand products is much more significant in percentage. The advantages the mean absolute percentage error brings are not needed in this research, which is why it is not used for this case. (Pascual, 2018)

2.5.3 Sample count

Forecast comparing is not accurate with only one product because there could be luck that some forecasting method is more suitable for this product, but it could be worse in general. There are also multiple kinds of products and customers in the food industry. That is why in this research, there are eight products chosen. These products should be different in many ways. (Godwin, 2021)

The age of the product is a significant difference, so there should be products that are old and young but not younger than one year so that the product has clear seasonality so that it is taken into account in this research. There should be products from grocery stores and food factories because it is assumed that food factories have more stable and more significant demand than grocery stores. The product demand size also affects forecast accuracy, so the research should include products with different demand sizes. These are the reasons why there should be eight products. (Godwin, 2021)

The products should be four from the grocery store. One with a short length of data and small sales, one old product with long sales history but small sales, one with a short length of data but large sales and one with a long sales history and many sales. For the food, factory research should use the same four kinds of products as the grocery store but from their perspective. (Godwin, 2021)

This means that there will be twenty products that all should forecast ten thirty-day periods. These forecasts make 6000 data points in total. These data points are made from a time period that already has a sale and is known by their demand. These 6000 forecast data points are compared to actual sales with the root mean square error calculation to see which forecast method is most accurate and suitable for the food industry. (Godwin, 2021)

3 Demand forecasting methods

This section presents seven different forecasting methods and some factors around demand forecasting. Forecasting methods chosen for testing are presented in the table belove.

Table 3. Tested forecasting methods, dataset features and testing parameters.

Method	Product count	Product dataset lengths	Error calculation	Test set length
Facebook Prophet	20	370 days - 1134 days	RMSE, RMSE %	30 days
LSTM	20	370 days - 1134 days	RMSE, RMSE %	30 days
CNN	20	370 days - 1134 days	RMSE, RMSE %	30 days
Average	20	30 days - 150 days	RMSE, RMSE %	30 days

Demand forecasting is the same as time-series forecasting. Time series is a series that contains values and time pairs. Forecasting time series is done by predicting future value for a specific time. Value is defined by looking into historical values. Many different calculations define the value in future. Usually, some software libraries contain these calculations. (Chatfield, 2000)

The following figure presents an example product from the dataset. From this figure, we can see that the data is highly seasonable. Another point from this figure is that data contains anomalies. These anomalies can cause significant errors in demand forecast accuracy.

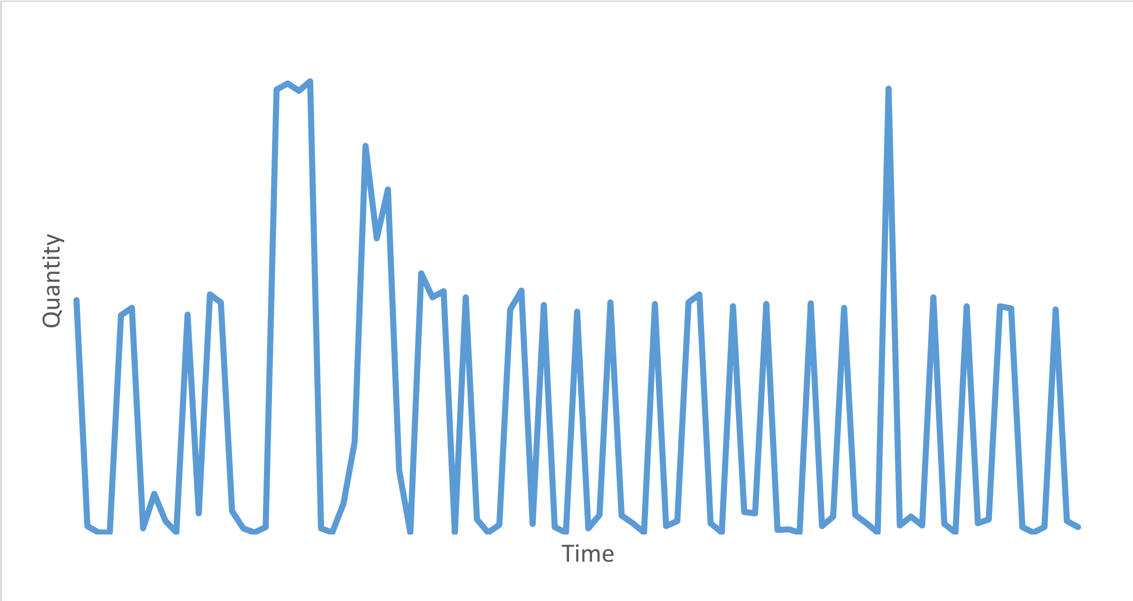

Figure 12. Example product from the test dataset.

3.1 Facebook Prophet

The Prophet is an Open-source project made by Facebook's Core Data Science team. Prophet analytics library is used for many of Facebook's use cases, and they have developed the library for their use because other libraries were not good enough. (Prophet, n.d.)

The Prophet is an analytics library that focuses on time series forecasting. Prophet tries to fit multiple seasonality non-linear trends to the data. Seasonality trends are fitted weekly, yearly, daily and for holidays. The Prophet makes fittings fully automatic but allows the user to tune the parameters and control the seasonality fittings with the parameters. Prophet can be found in python and R. (Prophet, n.d.)

Prophet supports a regression column, which is possible for forecasting campaigns by impaling the discount or product price to the regression column. The campaign should be added to the regression column in history and the future in these cases. Prophet fits the multiplier for the regression column component when campaigns are added to history. This multiplier is used in the future when there is a campaign. The regression column is linear, so more significant discount forecasts more significant sales as wanted. (Bakker, 2020)

Prophet also fits a holiday component for the forecast model. Finnish national holiday calendrer can be used for the holiday component, which is a significant advantage for the Prophet. Holidays have a significant effect on the demand of the food industry. (Guo, Ge, Jiang, Yao, & Hua, 2020)

For the food industry's needs, Prophet is a good fit for demand forecasting because it fits the model with different time periods, which can be found in the food industry demand. The food industry has a weekly seasonality component, and the demand is majorly different for weekday and weekend days. There is also different demand for some products at different times of the year, and Prophet can take these into account with the yearly seasonality component. (Guo, Ge, Jiang, Yao, & Hua, 2020)

Campaigns and Holidays also have significant effects on the food industry, and there is a component for these in Prophet. The Prophet considers many of the needed components in the food industry, so it is selected for testing. (Larson, 1997)

In this research, the Prophet is used with default parameters which means that some parameters are used with default constant values and for some, as seasonality Prophet selects it automatically depending on the data length. Python Prophet library also contains diagnostics tools for parameter adjustment, making it easy to implement and compare different parameters, but it is not included in this research. (Prophet, n.d.)

3.2 ARMA, ARIMA & SARIMA

ARMA is the shortening of the Autoregressive Moving Average. ARIMA and SARIMA are extending components for ARMA. ARIMA is the shortening of the Autoregressive Integrated Moving Average. SARIMA is the same as ARIMA, but it contains a seasonality component. (Pierre, 2021)

The base for these is the MA component which is the moving average. Moving average means calculating the average sale from a selected number of days. The number of days has a significant effect on the accuracy of the forecast. This presents two significant issues. It smoothens the changes inside the selected time period. If the count of the days is more extended than week/seven days, it will average the same value for the weekdays and weekend days. Special Events in demand would also be smoothed out. For example, a one-day event can be smoothened out before and after the event. So on the day of the event, there would need to be more products in stock and coming days after the event, the stock quantity would increase too high. Another area for improvement with the moving average is that it is slow in change and cannot anticipate changes. (corporatefinanceinstitute, 2022)

The problems presented with the moving average are trying to be solved with the Autoregressive component of the ARMA. The autoregressive component calculates connecting factors between the input and output data points. These data points can be in the sale history or the future, but they need to be known. This is how the autoregression added to the moving average can anticipate changes in the forecast. (Zhang & Moore, 2015)

SARIMA adds the seasonality component to the ARIMA. This is a significant advantage for the food industry, where seasonality affects demand. This makes the moving average model anticipate the seasonality changes in the demand. (Chen, Niu, Liu, Jiang, & Ma, 2018)

For the food industry, SARIMA is the best choice of these three, but the solution could be better. SARIMA is usually used for significant amounts of measurements, but it is not always the case in the food industry, and that is one disadvantage of SARIMA. For this research, the SARIMA is not included in the testing. (Chen, Niu, Liu, Jiang, & Ma, 2018)

3.3 CNN Convolutional Neural Network

CNN shorten of Convolutional neural network. The count of the dimensions in the CNN model defines different CNN solutions. There are one-, two- and three-dimension CNN models. (Verma, 2019)

One dimension values are, for example, time-series values, and in that, the dimension is the time. The time series values can have multiple data points at the same given time. In the food industry, the sales time series would be on input. However, it could also have different inputs such as campaigns, discounts, holidays or some other significant effects of that given time. It could solve many issues about these. It also gives significant possibilities for future regressors for the data because there is no limitation for affecting data points in the input. (Verma, 2019)

Another benefit of a one-dimension convolutional neural network is the built-in smoothing of the data. This also brings difficulty in how much sales history data could be smoothened to retain all essential features for the forecast. (Alex, 2020)

The two-dimension convolutional neural network is usually for image recognition. That is not feasible for time series forecasting, so it is not feasible for this research. (Yang, et al., 2018)

The three-dimensional convolutional neural networks are more challenging than the one and two-dimensional neural networks. Usually, it is used to process 3D images or videos. (Verma, 2019)

For these reasons, the one-dimension convolutional neural network is selected for testing in this research. This option would have many possibilities to use environmental knowledge to reach more accurate demand forecasts. However, it is not done in this research because it is done in a stable environment.

3.4 XGBoost

XGBoost is an open-source library for classification and regression problems. XGBoost distributed gradient-boosted decision tree machine learning library. This approach is used to solve many issues, such as automatic pricing for some items, houses, cars, and others. There is no limitation for the problem complexity, and input data can contain as many data points as needed. XGBoost is more effective than a traditional decision-making tree because it uses parallel decision-making trees, which we can see in the following architectural picture. The data is split into multiple subsets, and different decision-making trees handle those subsets.(Nvidia, n.d.)

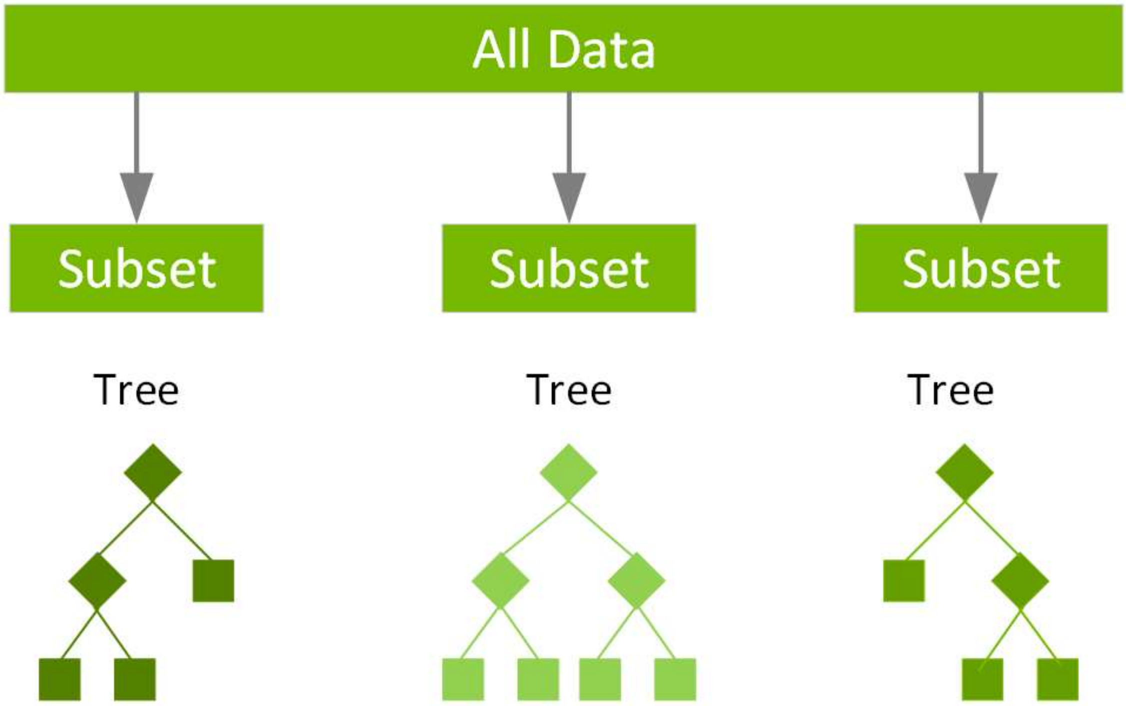

Figure 13. XGBoost parallel decision making tree architecture. (Nvidia, n.d.)

XGBoost can provide a significant advantage in the food industry when making large amounts of forecasts. XGBoost is a much different approach than other options in this research. The decision-making three needs to have a problem presented. This does not fit that well for time series forecasting as it is. For this reason, the sales history needs to be presented as the problem for the XGBoost and future demand as the solution for the problem. The decision-making three should use the sales history to train the model, which can be used after this for future demand forecasting. This approach allows the problem and, in this research case, the sales history and the environment to be more complicated than only one sale value at a given time. (Browlee, 2020)

Earlier research has studied XGBoost used on demand forecasting. Research done by Mohit, Yogesh, Prachi, Sandeep, Vijay and Sunil compares multiple methods for demand forecasting where one is XGBoost. They compared ARIMA, ARNN (Auto regressive neural network), SVM and XGBoost. They also included hybrid models, which combine multiple of these methods. (Gurnani, et al., 2017)

Their research found that when evaluating XGBoost against other individual methods with MAE, the XGBoost was the most accurate method, but when comparing it with RMSE, it was second. Also, hybrid models were generally more accurate than individual demand forecasting methods. (Gurnani, et al., 2017)

Existing research presents that hybrid models of XGBoost outperform the regular XGBoost. The research presents that combining two-step clustering with XGBoost would outperform XGBoost. This is called C-XGBoost. (Ji, Wang, Zhao, & Guo, 2019)

Other used hybrid model is A-XGBoost which combines ARIMA forecasting for linear parts and XGBoost for non-linear parts. This also outperforms XGBoost. (Ji, Wang, Zhao, & Guo, 2019)

Combining Clustering, ARIMA and XGBoost are also researched. This is called C-A-XGBoost. This has outperformed other hybrid models, as seen in the following table, which presents the results of the study that Shouwen, Xiaojing, Wenpeng and Dong have done. (Ji, Wang, Zhao, & Guo, 2019)

Table 4.XGBoost comparison to different hybrid model accuracy calculated with RMSE (Ji, Wang, Zhao, & Guo, 2019)

Error calculation method	XGBoost	A-XGBoost	C-XGBoost	C-A-XGBoost
RMSE	6,049	6,287	4,832	3,282

Because of its bad performance in earlier research with RMSE and that it is complicated to be used in demand forecasting for the problem-presenting approach and hybrid models, it is not included in this research for the study case.

3.5 Tensor Flow

Tensor flow is open-source machine learning and artificial intelligence library. Tensor flow is made by Google. It is a multipurpose library that implements deep neural networks. It can be used for various issues such as image recognition, text recognition, audio recognition and structured data. (J. R. S. Iruela, 2021)

The time series is structured data. For the time-series data, the Tensor Flow presents multiple solutions. The first one is the convolutional neural network. The second one is Recurrent Neural Networks and a linear solution. The third solution for demand forecasting included in Tensor Flow is Linear time series forecasting. These three options contract the model differently from what is seen in the following pictures. These different approaches are presented in the following architectural pictures. (TensorlFlow, 2022)

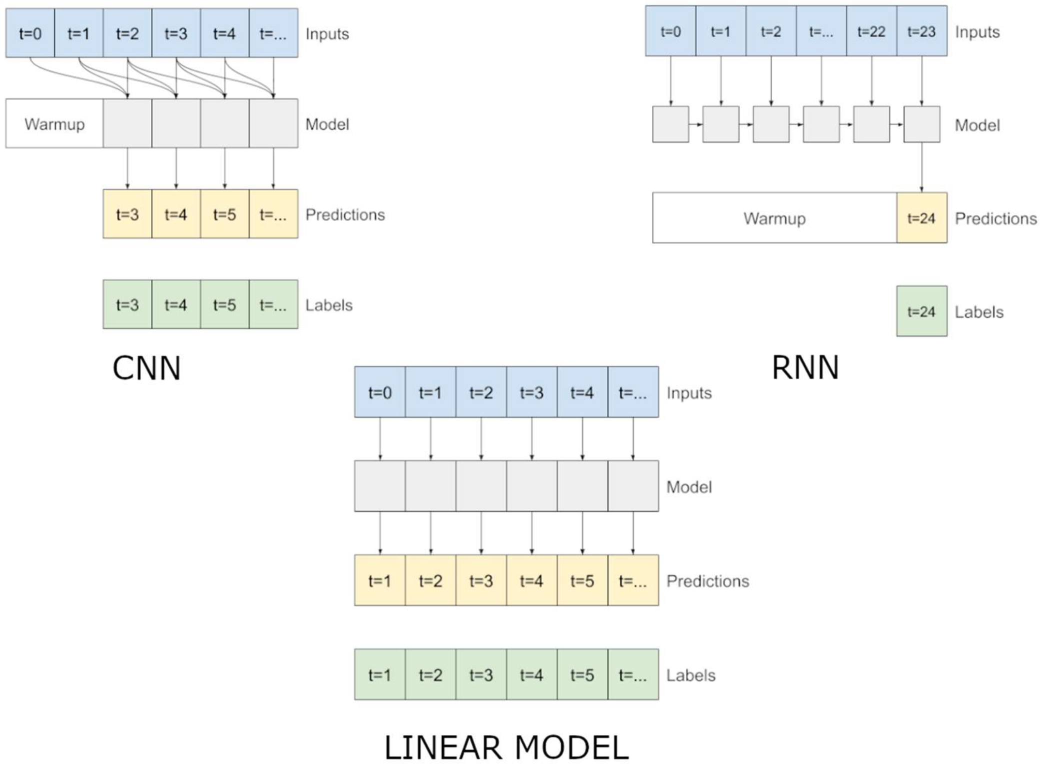

Figure 14. Architecture picture of CNN, RRN and Linear Tensor flow models (TensorlFlow, 2022)

These architecture diagrams show how the CNN combines the data. The RNN model combines the models for the forecasts. The linear model architecture shows its downside of using only the previous input to forecast the prediction. (TensorlFlow, 2022)

Existing research shows that also GRU (gated recurrent unit) can be used with Tensor Flow, and there are promising results of comparing Tensor Flow with GRU to LSTM. A study by Ivan, Noami, Nikica and Vinko presents that the accuracy for GRU was 81,3%, whereas, in the same study, it was only 77,6% for LSTM. This accuracy was found fresh goods dataset with CMAPE (corrected mean absolute percentage error). (Basljan, Munitic, Peric, & Lesic, 2021)

The convolutional neural network has been researched separately, so this part focuses on other options. A recurrent Neural Network is an excellent option for time series forecasting. The Recurrent Neural Network processes the data points step by step. The recurrent neural network does not present significant advantages for demand forecasting compared to the convolutional neural network. GRU presents promising results, but the studies done without squared error calculation do not value the uniformity of error and these questions the relevance of the earlier presented study. Therefore, Tensor Flow is not included in testing in this research. (TensorlFlow, 2022)

3.6 LSTM Long Short-Term Memory

LSTM is an abbreviation for Long Short-Term Memory. LSTM is a method of deep learning. LTSM is based on RNN. RNN stands for Recurrent Neural Networks. RNN contains an input layer, an output layer and multiple hidden layers. LSTM is structured similarly then RNN, but the difference is that there are feedback connections from the hidden layers to other hidden layers, as shown in the following architecture diagram. Therefore, LSTM can also find short-term patterns from the data. (Sankar, 2017)

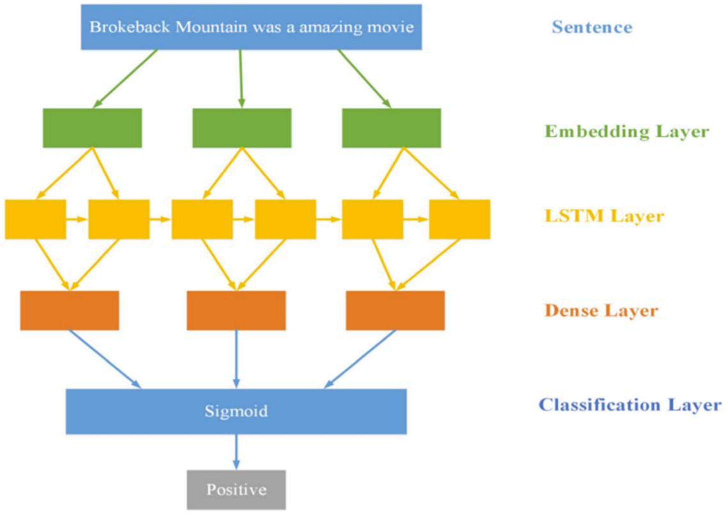

Figure 15. LSTM layer diagram about text classification research case. (Rehman, Malik, Raza, & Ali, 2019)

LSTM can be used for many different forecasts and predictions. For example:

- Solar radiation forecasting
- Time Series Prediction
- Fault prognosis of battery systems
- Earthquake trend prediction

(Mosavi, Ardabili, & Varkonyi-Koczy, 2020).

LSTM considers the seasonality changes in the time series dataset because it should be able to predict non-linear time series. LSTM has performed well in other research for time series forecasting. (Lindemann, Müller, Vietz, Jazdi, & Weyrich, 2021)

Earlier research compares the accuracy of LSTM and SARIMA models. This research does present that there are significant differences in results depending on the fore-casted product. (Falatouri, Darbanian, Brandtner, & Udokwu, 2022)

Table 5. Results table comparing LSTM and SARIMA in food products. (Falatouri, Darbanian, Brandtner, & Udokwu, 2022)

	RMSE		
Product	**LSTM**	**SARIMA**	**Most accurate**
Salad	949	1009	LSTM
Tomato	1170	1163	SARIMA
Potato	1184	1626	LSTM
Cucumber	4502	2854	SARIMA

As shown in the table above, there are no significant differences between SARIMA and LSTM. It depends on what earlier research has used that results in more accurate demand forecasts. (Falatouri, Darbanian, Brandtner, & Udokwu, 2022)

The LSTM is included in this research for testing mainly because of promising seasonality possibilities.

3.7 Average

Average is usually used with the safety stock calculation formula. Average is the traditional way to control and optimise stock quantities. The average length can be shorter than the total dataset length. The total dataset length is often too long to describe the present moment of the demand, and a moving average should be used instead. For the food industry, the average needs to be long enough to smooth out the weekend days, so it is essential that the length is the same as the total week count, so the average should be 7, 14, 21… For this test, it is chosen 14 days because it is the shortest option, which takes a more extended time period than one week, which is the length of shorter holidays in Finland. (Malato, 2020)

This research needs to include average calculation because it tells if existing stock quantity control methods are better than the demand forecasting one, which this master thesis suggests. This way average will be a great baseline when comparing forecasts.

3.8 Issues with demand forecasting

Demand forecasting should consider as many effects as possible, but it is not easy to take them into account because there is now historical data or information about these effects in the future. Many of these issues are excluded from this research, but they would need to be solved before the results of this research can be used in the food industry. Another possibility for using the results before solving the presented problems would be to use the demand forecasts in cooperation with humans to optimise the stock quantities. (Lutoslawski, et al., 2021)

3.8.1 Data collection

The base time series for demand forecasting is sales history. Sales history needs to contain at least the sales time and quantity. Sales history can also contain valuable information such as campaigns, prices or product placement. The information that affects product sales is valuable to understand why demand differs from average. Empty stock is also the essential information to understand why there are no sales for some period of time series. (Simplilearn, 2022)

In grocery stores, this data is collected to cashier systems which usually contain all needed information about sales history. The data is usually collected in the ERP system on the food production side. Data collection is easy to implement because it is usually collected in one place and can be fetched from one database. (Andarwati, Amrullah, Thamrin, & Muslikh, 2020)

Other valuable data can be used to enrich the sales history. In the food industry, holidays have a significant effect on demand. National holiday calendars are easy to get from calendar libraries. Future data can also contain valuable information which helps make the demand forecast. Available information can increase or decrease the demand forecast. The campaigns, price changes, holidays, and product placement are known for the future. Usually, future data can be collected from the same place as the historical sales data in the food industry. (Lewinson, 2022)

3.8.2 New product issues

New products do not have a sale history that could be used to make sale forecasts. Every other product has been new at some point, and humans can use those as reference values to see how new products usually affect sales of future new products. This should also be possible to calculate automatically, but the suggested methods do not take it into count. (Rogoza, 2019)

The average calculation would be easy to calculate from other products for the same day of the age of the product, but this would not be accurate because there are significant differences between different products in terms of the size of the demand. This could be solved by considering the products selling in similar counts to the new product. Again, this presents a new issue on how to forecast the seasonality, holidays and other known changes. (Rogoza, 2019)

New products often are also in campaigns or are marketed, and they might have different product placements. This causes issues when the campaigns and product placement changes because there is only sales data for the higher demand period because of the start marketing of new products. That is why the steady demand is still being determined and very difficult to forecast using this product's data. In these situations, the demand should be forecasted by seeing the demand change in other similar products for similar environmental changes because this complexity is not taken into account in this research. (Duffy, 1999)

3.8.3 Difference between sales history and demand history

Based on an accurate demand forecast should be accurate knowledge about the demand history. It is complicated to have data about the demand, so sales data is used as the base data of the demand forecast in this research. Often the sales data is equal to the demand, but there are exceptions. One exception where demand and sales differ is when the stock is empty. In these cases, the demand is higher than the sale. In minor differences, the accuracy of the demand forecast is not affected. However, in a more extended time period without sales, the product could be mistaken as a seasonality change and could be forecasted for the same time in the next year. (Yoon, 2021)

Another area for improvement in the sales data which can cause mistakes would be major sale events that would not duplicate in the future. These can be festivals, expositions and other events which happen only once or continuously at different times of the year. This event could be forecasted for the next year, which would cause much waste. (Marsh & Bugusu, 2007)

3.8.4 Campaigns

There are many effects on food demand. One significant effect is campaigns and discounts on the product. This is very difficult to take into account because the campaigns can change from 1 day to the total lifetime of the product. Different size discounts also affect differently. The discounts are often known in the historical sales data and for the future so that they can be considered. Campaigns usually do not affect only the price and the discount of the product but also the product and the discount are advertised to the customers. (Chinie, Biclesanu, & Bellini, 2021)

The market placement of the product can be changed in the store during the campaign, which significantly affects sales. It is not usually seen in the data and cannot be considered when making demand forecasts. These kinds of unknown changes can have a significant effect on the sale. Campaigns can also have limitations, such as buy three get one for free, which is also difficult to consider when making demand forecasts because there is no clear discount for the product, which could be compared to sales history. (Mummalaneni, Wang, Chintagunta, & Dhar, 2019)

3.8.5 Unusual demand changes

A significant issue is rapid changes in demand. These changes can be complex for the producer to anticipate. The changes can be temporary or permanent. This kind is caused, for example, by Covid, government regulation or unexpected publicity for the product. Usually, in these environments, the forecasting accuracy decreases, which can cause significant losses because the products could expire. Humans better know these changes, and in these cases, the human should be able to assist the forecasting method or how the results are used to avoid losses. (Wrang, 2022)

Another type of unusual change is slower changes in consumer habits. These can be significant changes in trends. One excellent example of these would-be consumers interest in vegan and organic products. These slower changes in demand trends should be considered in demand forecasting. (Wrang, 2022)

4 Results

The results section presents the demand methods compared against each other. First, the data set is presented. Next is the comparison of different baselines, which are different length averages. The most accurate average is selected from these, which is used as this study's baseline. The chosen methods Facebook Prophet, LSTM, and CNN, are implemented to forecast all products on the dataset. The demand forecasts accuracy is calculated for every forecasting method separately. After these results are presented for every forecasting method, they are compared against each other.

4.1 Dataset

Data used for this research is collected from grocery stores and food factories so that it widely presents the demand in the food industry. Accurate data is not presented in this research. However, as the following figure shows, the products forecasted in this research vary a lot from the smallest to the most significant products in demand.

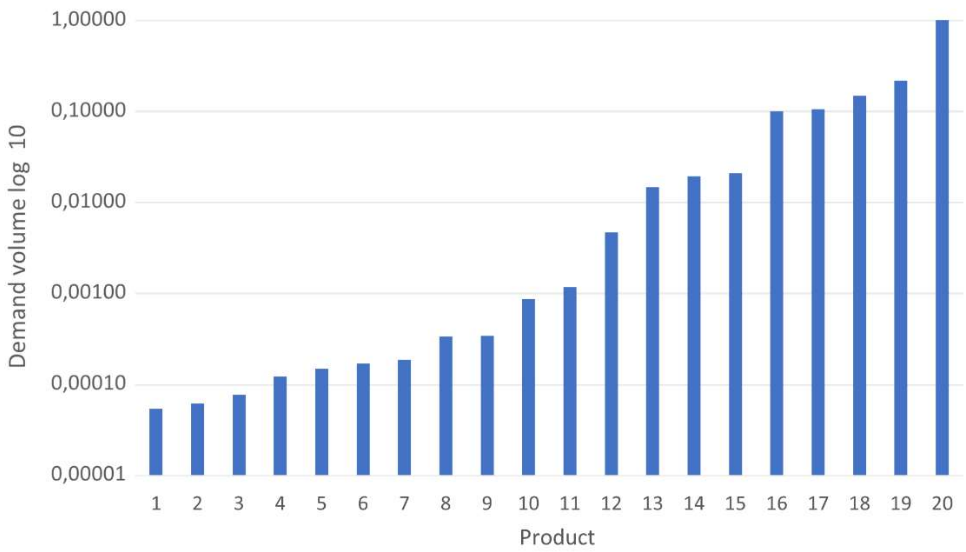

Figure 16 Dataset demand volume on logarithmic scale.

The previously presented dataset is scaled between 0 and 1. The volume differences are so significant that data volume axes are logarithmic so that all product volumes can be seen on the chart. This presents the volume differences in the dataset that correlate to volume differences in the food industry.

4.2 Average

Five different averages are used on the test data set in this research. The averages are five different lengths. Different lengths are 30 days, 60 days, 90 days, 120 days and 150 days. This research does not present all of them in detail. It selects the best one, presents that one in detail and compares it to other forecast methods.

4.2.1 Comparing different length averages

Different averages are compared with root mean square error.

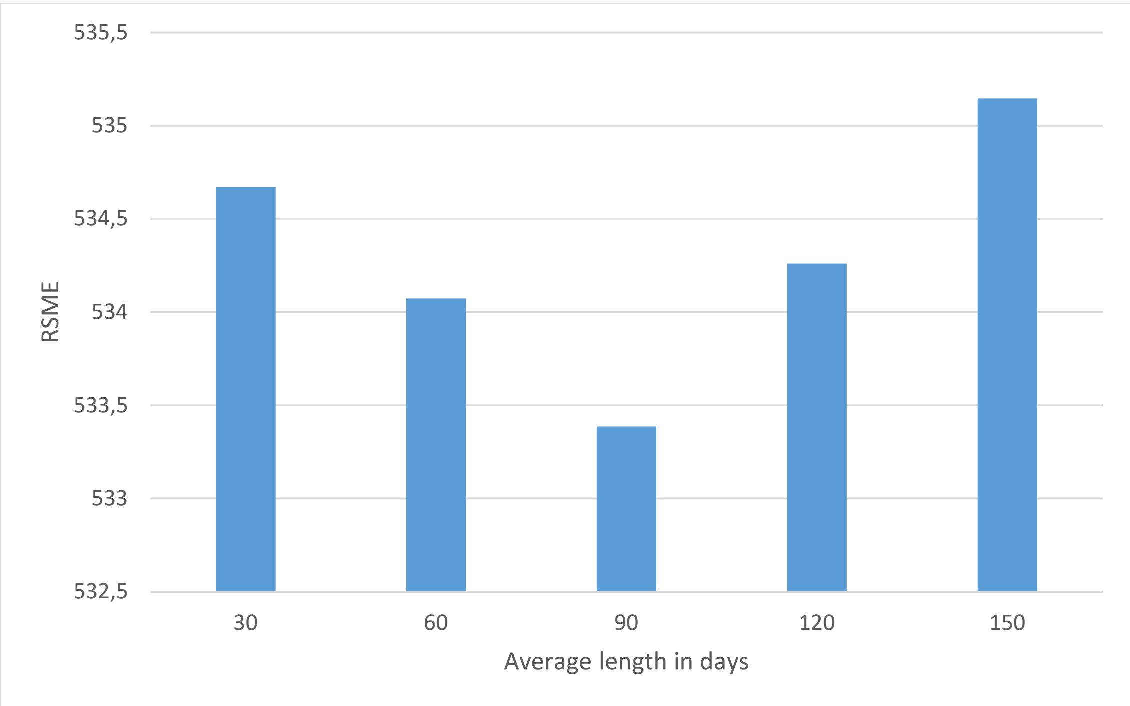

Figure 17. RMSE for different average lengths.

The figure above shows that 90 days average is the most accurate average length. Ninety days average is going to be the baseline which is compared to other forecast methods. The figure above also shows how much longer the average is and its loss of accuracy. The same happens with shorter average lengths. Ninety days is the most accurate option for this dataset, which presents the different demands of the food industry and is also suitable for different food industry sales data set.

The difference between the best average (90 days) and the worst average (150 days) is small. The most accurate RMSE is 533,4, and the worst is 535,1. That means there is a 0,3 per cent difference between best and worst accuracy. The per cent difference is small, but in the food industry, where the production volumes are high and amounts of capital assets invested into stocks are high, it might have a meaningful effect.

4.2.2 90 Days average

Comparing only the RMSE of the forecast highlights the error of large-volume products. Because the test set includes different volumes from a couple of product sales per day to thousands of days, it is essential to compare the error to the volume of the product. This is done by calculating the error percentage.

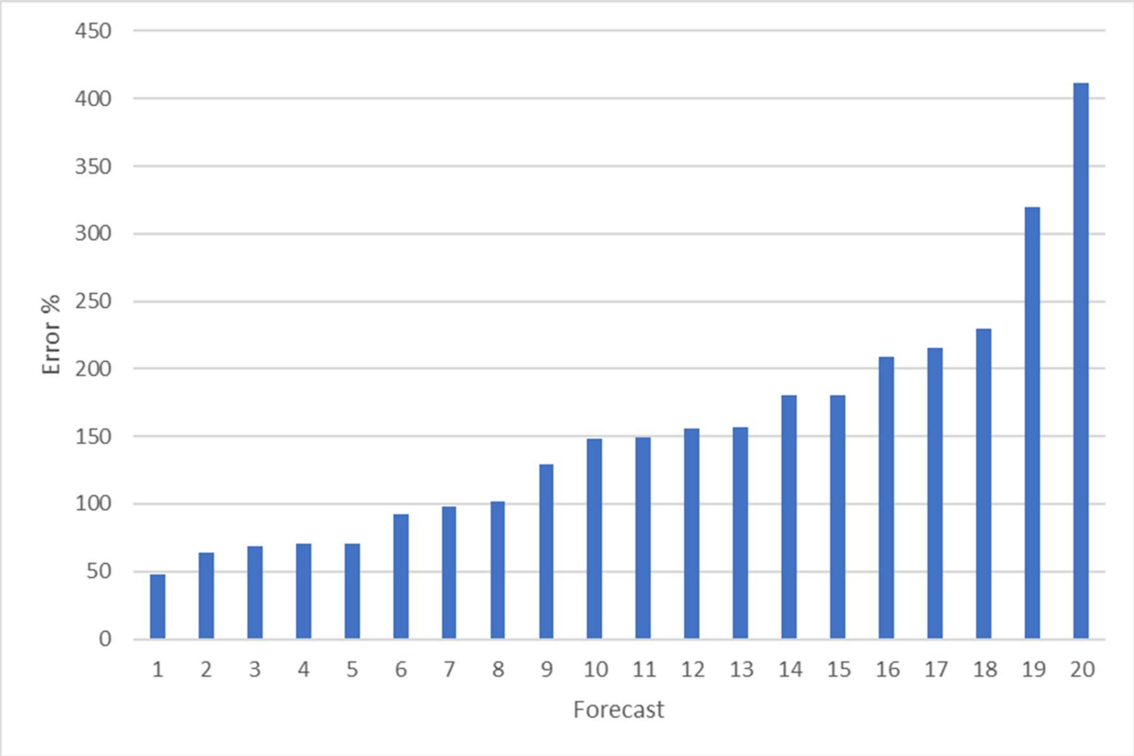

Figure 18. 20 Daily Averages RMSE error per cent.

The figure above shows the error percentage of daily RMSE values compared to daily average product sales. The error percentage for 90 days average is between 50 and 420 per cent. The average error percentage is 155 per cent. Daily accuracy is difficult because the data set contains products with significant differences between days. The daily error might not cause losses in the food industry because if the close-by days even the error out. This is why the same test has been done with weekly aggregated data.

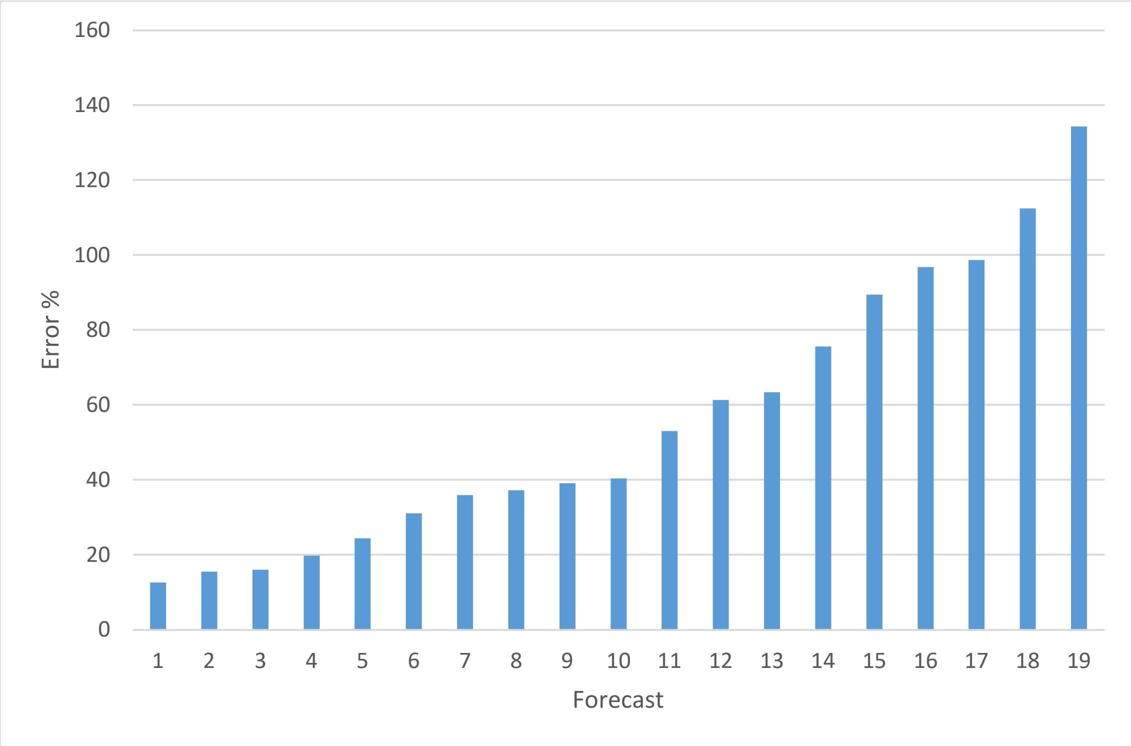

Figure 19. 20 Weekly Averages RMSE error per cent.

Error from the weekly aggregate data on the 90 days average is between 12 and 134 per cent, as we can see in Figure X. Average of this error per cent is 54 per cent. It shows how much better the average behaves weekly than daily.

4.3 Facebook Prophet

Facebook Prophet would have many parameters that could be adjusted, but in this research, the Facebook Prophet forecasting library was used with default parameters. The Facebook prophet is easy to define when using default parameters, which we can see in the following code snipped. Simply removing the test set for the training set, fitting the model, generating the future time window and making the prediction.

```
# Removing the test set from the trainging set

dataset = (dataset.loc[dataset['ds'] <= (dataset['ds'].max()
- timedelta(days=30))])
```

```
# Creating and fitting the model

model = Prophet()
model.fit(dataset)

# Defining the future timestamps

future = model.make_future_dataframe(
     periods= 30,
     freq='D',
     include_history=False)

# Forecasting for the future timestamps

result = model.predict(future)
```

Algorithm 1. Removing the test set from the dataset, fitting the dataset, generating the time window and making the prediction with Facebook Prophet.

Prophet's root mean square error is 402,3. Because this research did not run multiple parameter combinations for the demand forecasts, there is only one root mean square error value.

After the same error percentage calculation then for the average, the average daily error per cent for the Prophet forecasts is 168%.

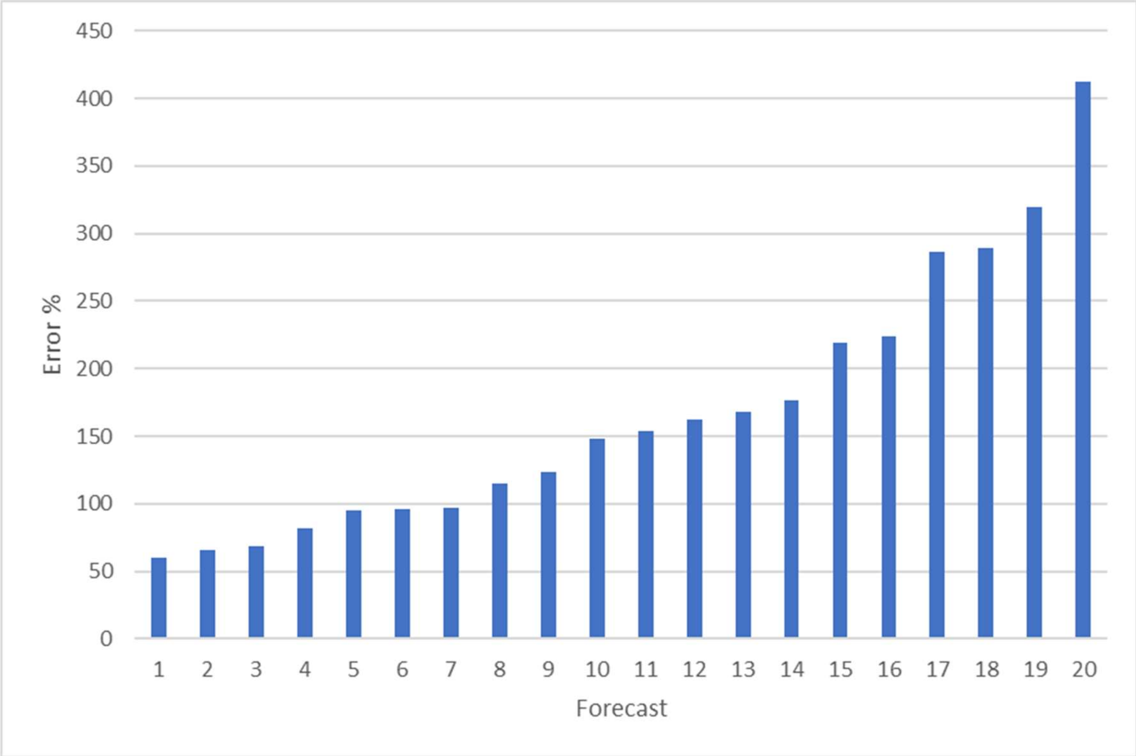

Figure 20. 20 Daily forecasts RMSE error per cent made with Facebook Prophet.

The figure above shows that the RMSE error percentage is between 55 and 415 per cent.

A similar weekly aggregated test for Prophet, then for average, tells us how accurate Prophet forecasts are on a weekly level. Weekly aggregate error for Prophet forecasts averages at 75 per cent.

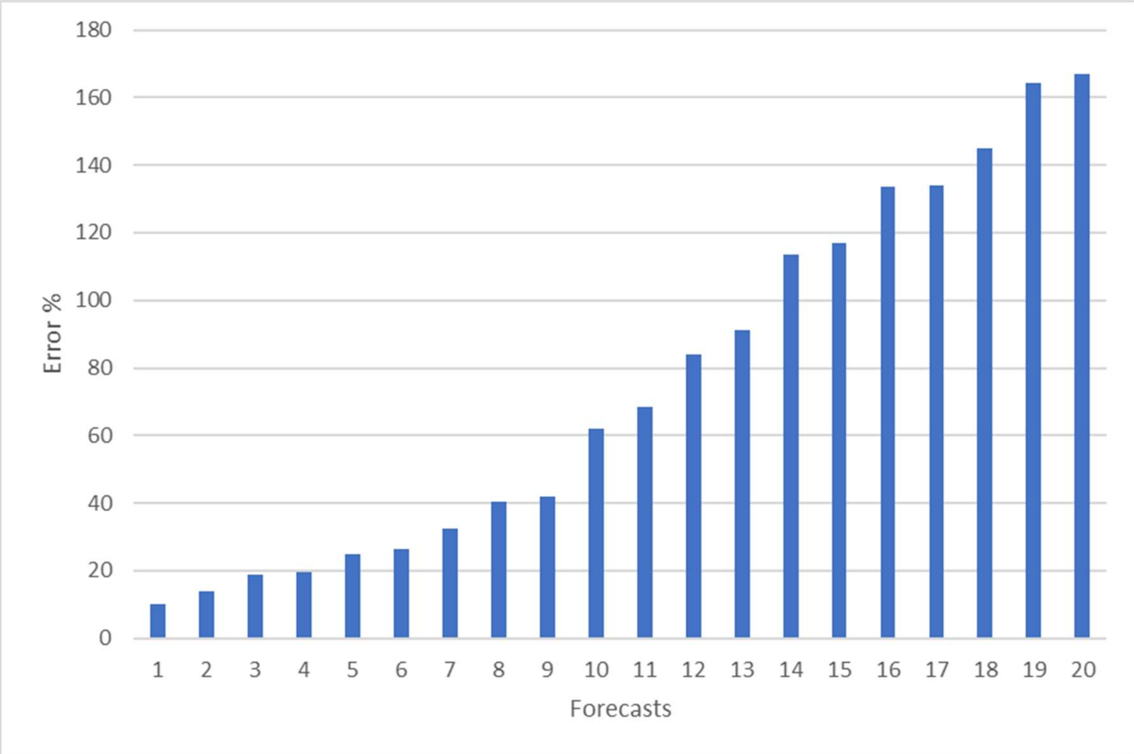

Figure 21. 20 Weekly forecasts RMSE error per cent made with Facebook Prophet.

Weekly forecast error is between 14 and 165 per cent.

4.4 LSTM

LSTM (Long Short-Term Memory) forecast has been done for the same dataset. For the LSTM, the dataset needs to be split into three parts. The training set, validation set and test set. The test set is the same time period as for other forecast methods. It is the last 30 days of the data set. The remaining part of the data set is split into training and validation sets. The training set is 75 per cent from the remaining dataset, and the validation set is 25 per cent from the remaining dataset, as we can see in the following code snipped. LSTM fits the training set as well as possible and validates this with the validation dataset. The validation error calculation method was chosen to be RMSE. IT was chosen because it is the same error calculation method used later in this research to compare different forecast methods.

```
# Define the training, validation and test set

training = dataset[:int((len(dataset) - WINDOW_SIZE) *
0.75)]
validation = dataset[int((len(dataset) -WINDOW_SIZE) *
0.75):(len(df)-(WINDOW_SIZE))]
test = dataset[(len(dataset) - (WINDOW_SIZE)):]

# Define the forecast window

wg = WindowGenerator(
     input_width=WINDOW_SIZE+1,
     shift=1,
     label_width=WINDOW_SIZE,
     label_columns=['quantity'],
     train_df=training,
     val_df=validation,
     test_df=test)

# Define and fit LSTM

model = keras.Sequential([
     tf.keras.layers.LSTM(256, return_sequences=True),
     tf.keras.layers.Dense(units=1)],
     name='lstm')

model.compile(
     optimizer=tf.keras.optimizers.Adam(learning_rate=0.000
     1),
     loss=tf.keras.losses.MeanSquaredError(),
     metrics=[RootMeanSquaredError()]
model.fit(
     wg.training_dataset,
     validation_data=wg.validation_dataset,
          epochs=200,
          verbose=2,
          callbacks=[tf.keras.callbacks.EarlyStopping(monit
      or='val_loss',
               patience=2,
               mode='min')])

# Make prediction

result = wg.predict(test, model)
```

Algorithm 2. LSTM training, validation and test set definition. Time window generation and model fit.

LSTM forecasts error calculation with RMSE averages of 727 on the 20 different products of the dataset. A similar error per cent calculation has been done for the forecasts from LSTM. The average RMSE percentage error is 168 per cent. From the figure belove, we can see that the error percentage is between 45 and 421.

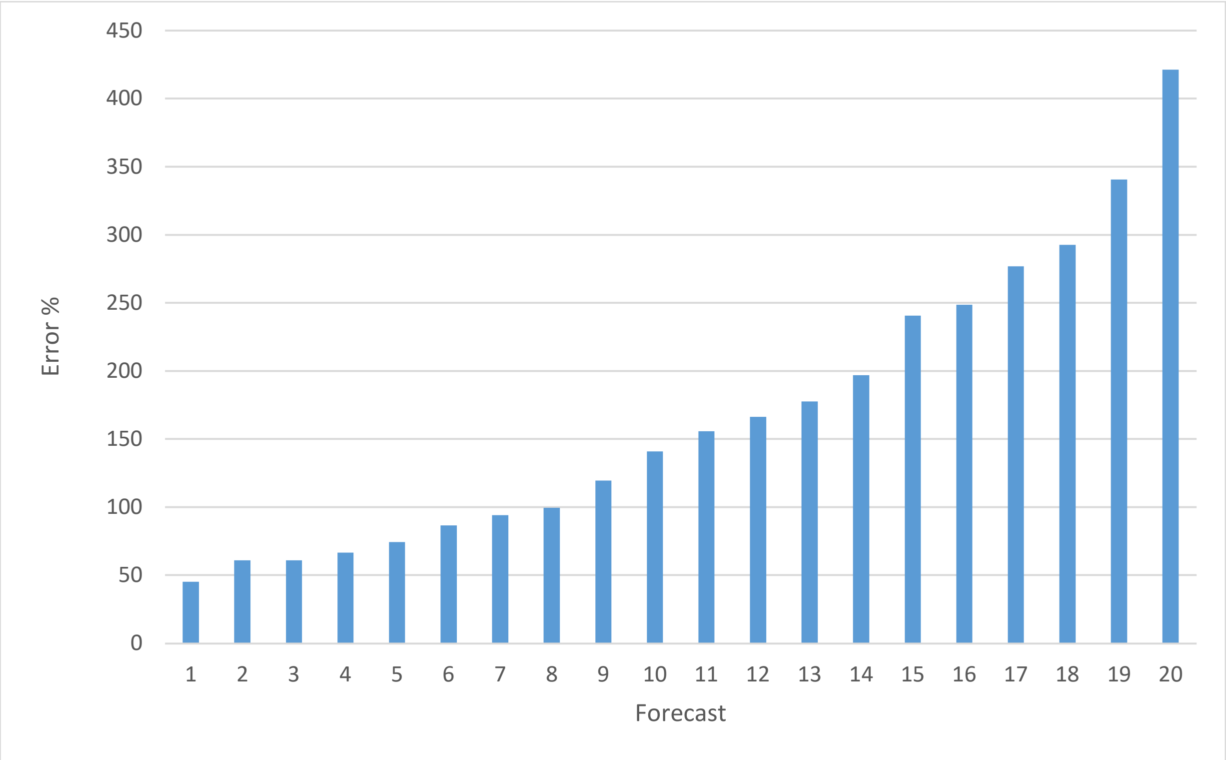

Figure 22.Daily forecasts RMSE error per cent made with LSTM.

Data is aggregated weekly, and the same error calculation is performed on weekly aggregated data. Weekly aggregated data gives RMSE percentages for every product, which can be seen in the following figure. The result from this study shows that the RMSE average error per cent of LSTM forecasts for weekly aggregated data is 77 per cent.

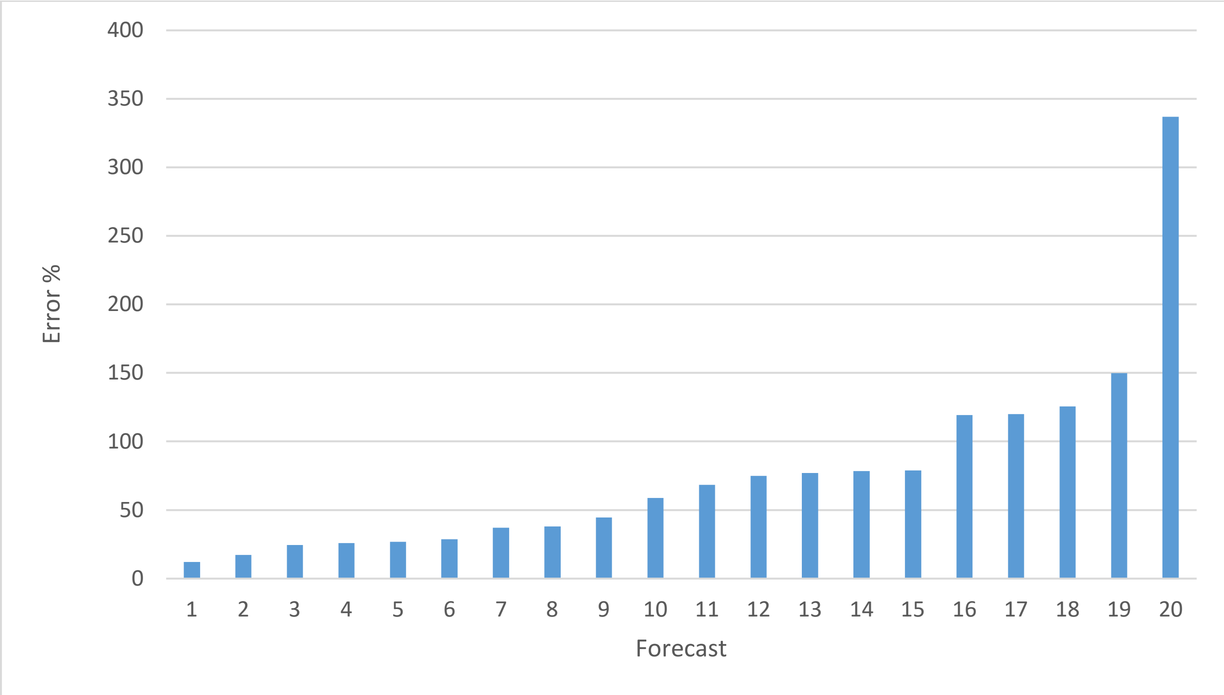

Figure 23. Weekly forecasts RMSE error per cent made with LSTM.

The figure above shows that the weekly aggregate data RMSE error percentage is between 12 and 336. We can also see a significant spike in the data with the worst product of forecasts. Otherwise, LSTM is performing well.

It can also be seen from the figure that the accuracy between the most inaccurate and second inaccurate forecasts is over double. The most minor volume product causes it. With a very small volume, the error percentage is high even with a small error in quantity.

4.5 CNN

Fitting the CNN model with a similar data set split than with LSTM. The fitting dataset is 75 per cent and 25 per cent for the validations set. The test set is performed with the same dataset as for every other forecast method. The average daily RMSE for CNN forecast results is 646. The same Daily and weekly error calculations as for other forecasting methods are performed for CNN.

```
# Define the training, validation and test set

training = dataset[:int((len(dataset) - WINDOW_SIZE) *
0.75)]
validation = dataset[int((len(dataset) -WINDOW_SIZE) *
0.75):(len(df)-(WINDOW_SIZE))]
test = dataset[(len(dataset) - (WINDOW_SIZE)):]

# Define the forecast window

wg = WindowGenerator(
     input_width=WINDOW_SIZE+1,
     shift=1,
     label_width=WINDOW_SIZE,
     label_columns=['quantity'],
     train_df=training,
     val_df=validation,
     test_df=test)

# Define and fit CNN

model = keras.Sequential([
     tf.keras.layers.Conv1D(64, kernel_size=2),
     tf.keras.layers.Dense(units=1)],
     name='cnn')

model.compile(
     optimizer=tf.keras.optimizers.Adam(learning_rate=0.000
     1),
     loss=tf.keras.losses.MeanSquaredError(),
     metrics=[RootMeanSquaredError()])

model.fit(
     wg.training_dataset,
     validation_data=wg.validation_dataset,
     epochs=200,
     verbose=2,
     callbacks=[tf.keras.callbacks.EarlyStopping(
          monitor='val_loss',
          patience=2,
          mode='min')])

# Make prediction

result = wg.predict(test, model)
```

Algorithm 3. CNN training, validation and test set definition. Time window generation and model fit.

CNN daily RMSE average error per cent is 159 per cent. As we can see from the following figure, the error is between 41 and 422 per cent.

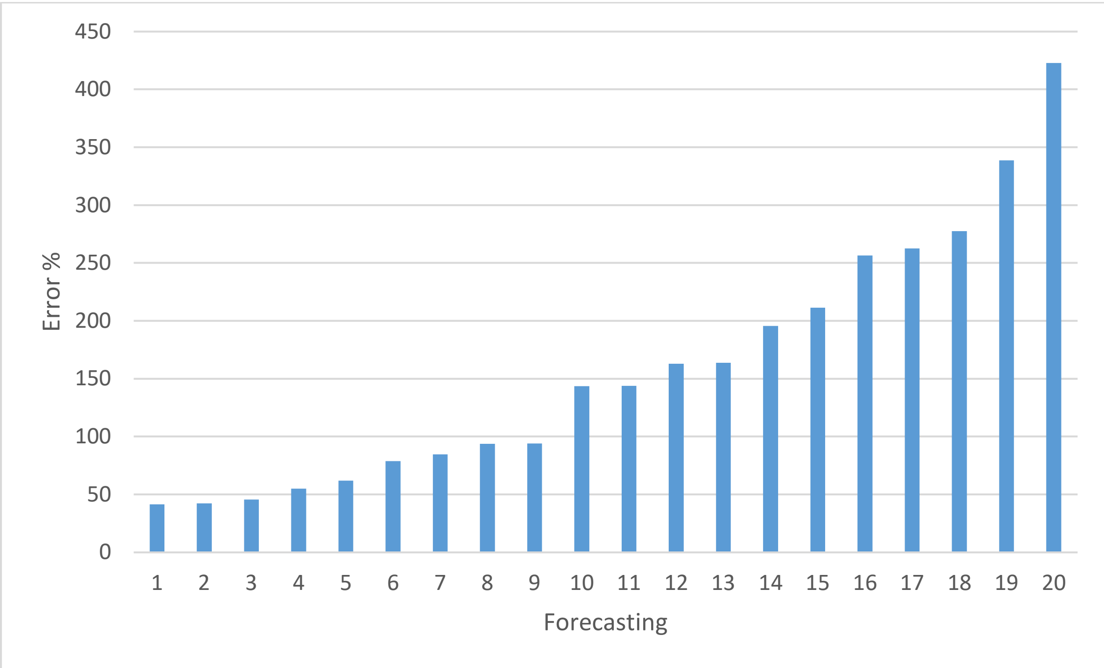

Figure 24. Daily forecasts RMSE error per cent made with CNN.

Weekly aggregated error calculation for CNN can be seen in figure x. The weekly aggregate error calculation average is 78 per cent. At the lowest, the error is 7 %, and at the highest, it is 418 per cent. It is seen in the figure that one forecast test has a lot higher error than others. The second highest error is 143 per cent. It is the most significant difference between the two forecasts in the research. The difference is over two times the second forecast.

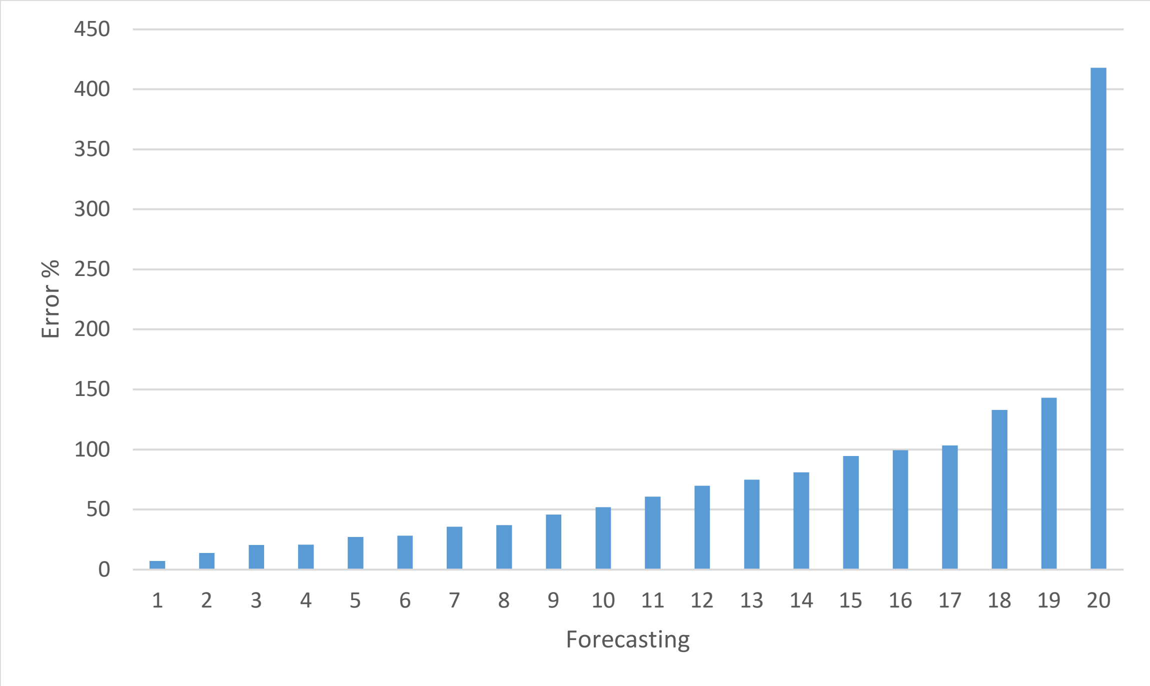

Figure 25 Daily forecasts RMSE error per cent made with CNN.

4.6 Comparing methods

Earlier separately tested forecasting results are compared in this section. Differences in daily error per cent are smaller. Error percentage varies between 168,25 and 154,92, per cent. The worst daily error per cent results from the LSTM forecasting test. The most accurate forecasting test was with 90 days average, as we can see in the following table.

Table 6. Error per centages and accuracy order in daily and weekly aggregations.

	Daily	Weekly	Most accurate Daily	Most accurate weekly
LSTM	168,25 %	77,11 %	3	4
CNN	158,82 %	78,25 %	4	2
Prophet	168,05 %	75,44 %	2	3
90 Average	154,92 %	53,99 %	1	1

Surprisingly, the average gives the best results for Daily error. The average is the same for every day of the test time period, and it does not take into account the difference between weekdays and weekend days. Every other method is considering the changes inside a week and trying to forecast them. For these reasons, it is weird that why other methods are worse on a daily level. One reason the average been the most accurate in daily comparison could be that any of the methods are not limited to only forecasting positive values. The average never forecasts negative values because the input values are positive or zero. Comparing to other methods, which can forecast negative values in situations where the trend is going down. From other methods, then average, the LSTM and Prophet are close to equal. CNN is the second most accurate on the daily comparison. It is only 4 per cent away from the most accurate, as shown in the following figure.

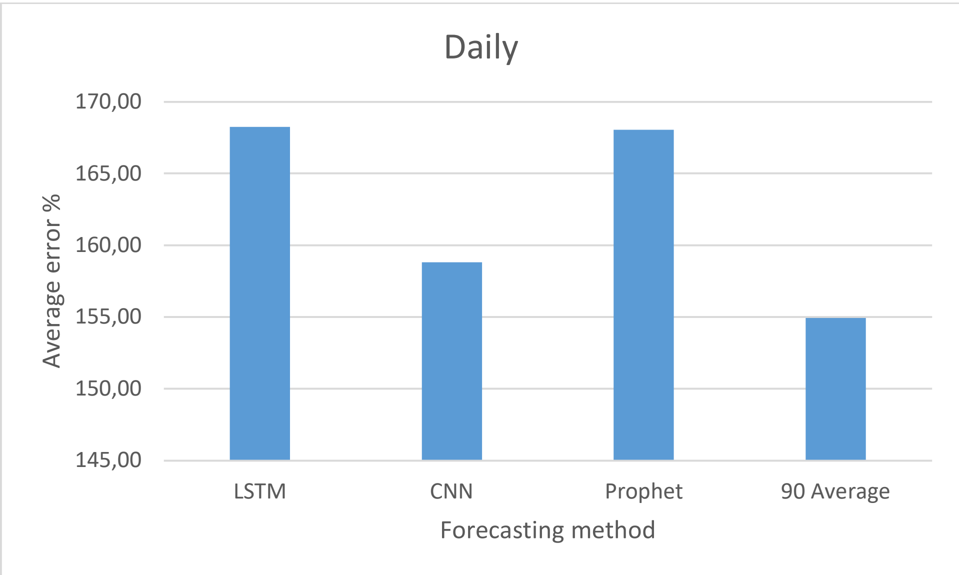

Figure 26 Daily forecasting method comparison.

Weekly comparison error percentage varies between 78,25 and 53,99 per cent. Error per cent is much smaller than with daily and this was presumable because the sums of weeks are higher than daily sales. The most accurate weekly forecast is 90 days on average. It is not surprising knowing that it was most accurate the daily result.

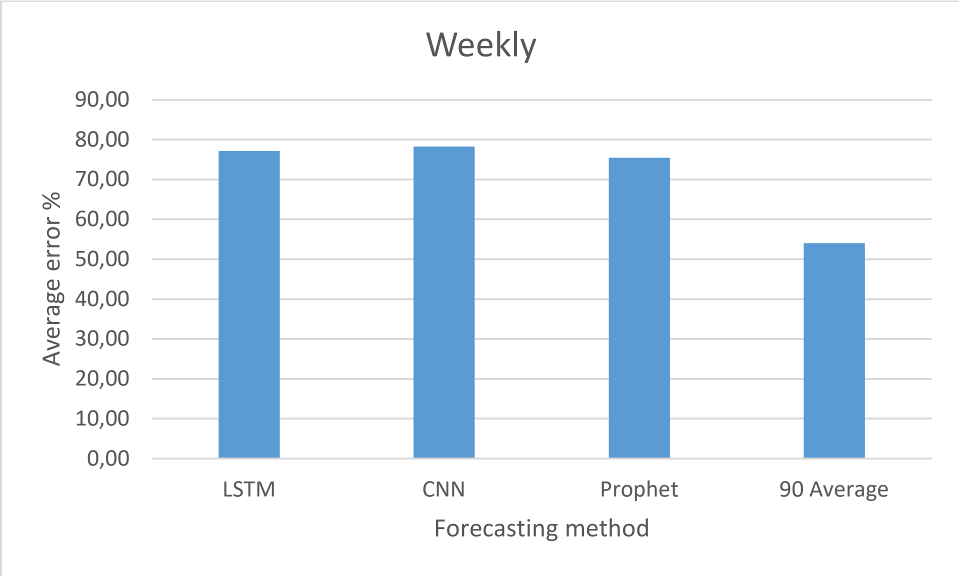

Figure 27 Weekly comparison of different forecasting methods.

Other methods than 90 days average are close to each other and around 20 to 25 per cent worse than 90 days average. During the forecast time period, other methods than average are adjusting for the trend of the product. These methods also adjust for yearly seasonality, which should give them advance when comparing them against the average. That is why it is surprising that 90 days average is the most accurate method.

LSTM, CNN and Prophet are under 3 per cent of each other. That tells us it could be assumed that the methods are taking trends and seasonality into account similarity. The unusual demand changes have been discussed earlier in this research. However, it is essential to note that the sales data might have different trends and seasonality than the couple last years because of the changes in demand caused by the Coronavirus pandemic during the years after 2020. During unstable times, the average has significantly advanced compared to methods that try to correlate the seasonality and trend changes from earlier years.

5 Conclusion

The answer to the research question “what is the suitable way for the food industry to make demand forecasting?” is a surprisingly average calculation. The average calculation was the research baseline. There are also differences between different length average calculations, which were included in this research. Ninety days was the most accurate average and most accurate forecasting method.

This is surprising, taking into account other research that has better accuracy with other methods than average. Research also so that the food industry has strong seasonality and product trends, which should give a significant advantage for different methods than average. The difference between weekdays and weekend days is shown by other research, and this should have given a significant advantage for other methods than average. For these reasons, this research presents a question of why the average is more accurate than other methods.

Compared to previous research, some results are conflicted because previous research has found that it is more accurate to use a method that considers seasonality than a method that does not. For example, Abbasimehr and Nahari have found this in their research. (Abbasimehr & Nahari, 2020)

Another research has found similar results as this research. Research made by Yunishafira is made using demand data from the garment industry, but it is researching demand forecasting and found that average is the most accurate way to forecast demand in the garment industry. (Yunishafira, 2018)

Results present that choosing the correct method in the food industry is essential, and this should be considered when using forecasts for stock quantity optimisation. Average is often much easier to implement the other forecasts, and it could be the first choice when building the infrastructure for restocking systems in the food industry. If there are available resources to research the other options for the system, that should be done. Because this research does not take into account many significant known changes in the food industry demand as holidays, events etc. Other things this research does not consider are the different sizes in demand for products. There could be different results if this research did the testing in separate groups for different volume products. The main points from this research for a person in the food industry are:

- Use data-driven decision-making to lower waste, cost and unnecessary capital assets used on stocks.
- Use average calculation if there are no resources to research and implement a more complicated system, the default solution.
- Understand the structure of demand as seasonality at yearly and weekly levels. Use this understanding to choose the approach for demand forecasting.
- Knowing the error of using the demand forecast method and using that to adjust the stocks without losing sales.
- That it would be possible to use the results of this research and demand forecasting, the base work should be done before that. Implementing entity should first choose the correct inventory model for their requirements. They should look into the benefits of different models or test them in practice.

The benefits of other forecasting methods should be remembered. Existing research shows that there is much greater potential when demand forecasting with other methods than average. Seasonality significantly affects the food industry demand, and other methods can take this into account. Facebook Prophet could be easily implemented to consider seasonality, which is why someone implementing demand forecasting in the food industry use case should look into using this kind of method rather than using average, even if it were more accurate in this research case.

All other forecasting methods have lots of parameters and ways to which could be finetuned than the average. This creates an excellent opportunity for future research on how finetuning would affect the accuracy of forecasting methods. The forecast methods also can include other factors into forecasts such as holidays, discounts, lack of stock and any other numeric or boolean value that could be associated with the sales. Taking these features into use for the forecasting method means that they should also be included in the testing dataset, which this research did leave outside by purpose. One major disadvantage of this research is that all methods have predicted demand separately by only using the data from one-time series. All 20 products used in this research are from the same environment, and there could be possibilities to use all of this to create a more accurate demand forecast for every product. The demand forecasting methods section of this research shows that there are multiple different methods to do demand forecasting, which this research did not include. Including these methods could present an excellent base for future research.

References

A. Olsson, C. S. (2008). *Risk Management and Quality Assurance Through the Food Supply*, 52-53.

Abbasimehr, H.;& Nahari, M. K. (2020). Improving Demand Forecasting with LSTM by Taking into Account the Seasonality of Data. 187.

Adebanjo, D.;& Mann, R. (2000). Identifying problems in forecasting consumer demand in the fast moving consumer goods sector. 225.

Aijaz, I.;& Agarwal, P. (2019). *A Study on Time Series Forecasting using Hybridization of Time Series*, 2-4.

Alex. (14. April 2020). *Boostedml*. Noudettu osoitteesta 1-d Convolutional Neural Networks for Time Series: Basic Intuition: https://boostedml.com/2020/04/1-d-convolutional-neural-networks-for-time-series-basic-intuition.html

Alexander, P.;Brown, C.;Arneth, A.;Finnigan, J.;Moran, D.;& Rounsevell, M. (2017). *Losses, inefficiencies and waste in the global food system*, 1-2.

Andarwati, M.;Amrullah, F.;Thamrin, E.;& Muslikh, A. R. (2020). *An Analysis of Point of Sales (POS) Information Systems in*, 1-3.

Bakker, L. A. (2020, June 9). Retrieved from Business forecasting with Facebook Prophet

Basljan, I.;Munitic, N. F.;Peric, N.;& Lesic, V. (2021). Prediction of perishable goods deliveries by GRU neural networks for reduction of logistics costs. *IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)*, 1-6.

Berry, S.;& Haile, P. (2021). FOUNDATIONS OF DEMAND ESTIMATION. 3.

Browlee, J. (5. August 2020). *Machine Learning Mastery*. Noudettu osoitteesta How to Use XGBoost for Time Series Forecasting: https://machinelearningmastery.com/xgboost-for-time-series-forecasting/

Chatfield, C. (2000). *TIME-SERIES FORECASTING.* Boca Roaton: Chapman & hall/CRC.

Chen, F.;& Ou, T. (2008). A Neural-Network-Based Forecasting Method for Ordering Perishable Food. *Fourth International Conference on Natural Computation*, 250-254.

Chen, P.;Niu, A.;Liu, D.;Jiang, W.;& Ma, B. (2018). Time Series Forecasting of Temperatures using SARIMA: An. 1-3.

Chinie, C.;Biclesanu, I.;& Bellini, F. (2021). *The Impact of Awareness Campaigns on Combating the Food*, 6-9.

corporatefinanceinstitute. (2022). Noudettu osoitteesta Moving Average.

Du, X. F.;Leung, S. C.;Zhang, J. L.;& Lai, K. K. (2011). Demand forecasting of perishable farm products using support vector machine. *International Journal of Systems Science*, 556-565.

Duffy, M. (1999). *The influence of advertising on the pattern of food consumption in the UK*.

Falatouri, T.;Darbanian, F.;Brandtner, P.;& Udokwu, C. (2022). Predictive Analytics for Demand Forecasting – A Comparison of SARIMA and LSTM in Retail SCM. *3rd International Conference on Industry 4.0 and Smart Manufacturing*, 994-1003.

Fildes, R.;Ma, S.;& Kolassa, S. (2019). Retail forecasting: Research and practice. *International Journal of Forecasting*, 11-13.

Folorunso, O.;Adewale , G.;Adewale , O. O.;& Okesola , O. J. (2011). Pinch Analysis as a Knowledge Management Tool for Optimization in Supply Chain. 80-83.

Godwin, J. A. (14. June 2021). *Towards Data Science*. Noudettu osoitteesta Time Series Analysis: https://towardsdatascience.com/time-series-analysis-7138ec68754a

Guo, C.;Ge, Q.;Jiang, H.;Yao, G.;& Hua, Q. (2020). *Maximum Power Demand Prediction Using*, 5-6.

Gurnani, M.;Korke, Y.;Shahz, P.;Udmale, S.;Sambhe, V.;& Bhirudk, S. (2017). Forecasting of sales by using fusion of Machine Learning techniques. *International Conference on Data Management, Analytics and Innovation (ICDMAI)* , 93-101.

Huber, J.;Gossmann, A.;& Stuckenschmidt, H. (2017). Cluster-based hierarchical demand forecasting for perishable goods. *Expert Systems with Applications*, 144-150.

J. R. S. Iruela, L. G. (2021). *A TensorFlow Approach to Data Analysis for Time Series*, 8.

Ji, S.;Wang, X.;Zhao, W.;& Guo, D. (2019). An Application of a Three-Stage XGBoost-Based Model to Sales Forecasting of a Cross-Border E-Commerce Enterprise. *Mathematical Problems in Engineering*, 1-15.

Khair, U.;Fahmi, H.;Hakim, S.;& Rahim, R. (2002). *Forecasting Error Calculation with Mean Absolute Deviation*, 1-5.

King, P. L. (2011). *Understanding safety stock and mastering its equations*, 1-3.

Kiuru, S.;& Silvennoinen, K. (2016). *Luonnonvarakeskus*. Noudettu osoitteesta https://projects.luke.fi/ravintolafoorumi/ruokahavikin-maara-laatu-ravitsemispalveluissa/

Koponen, P.;Ikäheimo, J.;Koskela, J.;Brester, C.;& Niska, H. (2020). *Assessing and Comparing Short Term Load*, 5-6.

Larson, R. (1997). *Food Consumption and Seasonality*, 37-41.

Lewinson, E. (16. January 2022). *Towards Data Science*. Noudettu osoitteesta The Easiest Way to Identify Holidays in Python: https://towardsdatascience.com/the-easiest-way-to-identify-holidays-in-python-58333176af4f

Lindemann, B.;Müller, T.;Vietz, H.;Jazdi, N.;& Weyrich, M. (2021). A survey on long short-term memory networks for time series prediction. 653-655.

Lisan, S. (2018). *Safety Stock Determination of Uncertain Demand and*, 5-8.

Lutoslawski, K.;Hernes, M.;Radomska, J.;Hajdas, M.;Walaszczyk, E.;& Kozina, A. (2021). *Food Demand Prediction Using the Nonlinear*, 1-3.

Malato, G. (23. June 2020). *Towards Data Science*. Noudettu osoitteesta An algorithm to find the best moving average for stock trading: https://towardsdatascience.com/an-algorithm-to-find-the-best-moving-average-for-stock-trading-1b024672299c

Marsh, K.;& Bugusu, B. (2007). *Food Packaging—Roles, Materials,and Environmental Issues*, 1-4.

Mosavi, A.;Ardabili, S.;& Varkonyi-Koczy, A. (2020). List of Deep Learning Models. 208.

Mummalaneni, S.;Wang, Y.;Chintagunta, P.;& Dhar, S. (2019). *Product Placement Effects on Store Sales: Evidence*, 2-6.

N. de P. Barbosa, E. S. (2015). DEMAND FORECASTING FOR PRODUCTION PLANNING IN. 3.

Nvidia. (ei pvm). Noudettu osoitteesta XGBOOST: https://www.nvidia.com/en-us/glossary/data-science/xgboost/

Pascual, C. (16. September 2018). *Dataquest*. Noudettu osoitteesta Tutorial: Understanding Regression Error Metrics in Python: https://www.dataquest.io/blog/understanding-regression-error-metrics/

Pierre, S. (5. October 2021). *A Guide to Time Series Forecasting in Python*. Noudettu osoitteesta https://builtin.com/data-science/time-series-forecasting-python

Priymasiwi, M. H. (2018). *Optimization Of Raw Material Inventory Costs In The Food*, 1-2.

Prophet, F. (ei pvm). *Prophet*. Noudettu osoitteesta Prophet: https://facebook.github.io/prophet/

Rehman, A. U.;Malik, A. K.;Raza, B.;& Ali, W. (2019). A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis. *Multimedia Tools and Applications*, 10.

Rogoza, W. (2019). *Method for the prediction of time series using small sets of*, 1-4.

Sankar, S. (2017). Recent Advances in Recurrent Neural Networks. 1-10.

Silva, J.;Figueiredo, M.;& Braga, A. (2019). *Demand Forecasting: A Case Study in the Food*, 5.

Simplilearn. (3. March 2022). Noudettu osoitteesta What Is Data Collection: Methods, Types, Tools, and Techniques: https://www.simplilearn.com/what-is-data-collection-article

TensorlFlow. (26. January 2022). Noudettu osoitteesta Time series forecasting: https://www.tensorflow.org/tutorials/structured_data/time_series

Van Donselaar, K. H.;Peters, J.;De Jong, A. H.;& Broekmeulen, R. (2012). Analysis and forecasting of demand during promotions for perishable items. *International Journal of Production Economics*, 65-75.

Verma, S. (20. September 2019). *Towards Data Science*. Noudettu osoitteesta Understanding 1D and 3D Convolution Neural Network | Keras:

https://towardsdatascience.com/understanding-1d-and-3d-convolution-neural-network-keras-9d8f76e29610

Wang, J.;Liu, G.;& Liu, L. (2019). IEEE International Conference on Big Data Analytics. *A Selection of Advanced Technologies for Demand Forecasting in the Retail*, 317-319.

Willems, S. P. (2001). Inventory Basics.

Wrang, E. (9. April 2022). *Bussiness Finland*. Noudettu osoitteesta A NEWLY COMPLETED STRATEGY POINTS THE WAY TO THE FUTURE OF FOOD EXPORTS: https://www.businessfinland.fi/en/whats-new/blogs/2022/a-newly-completed-strategy-points-the-way-to-the-future-of-food-exports

Yang, X.;Ye, Y.;Li, X.;Lau, R.;Zhang, X.;& Huang, X. (2018). *Hyperspectral Image Classification With Deep Learning Models*, 1-5.

Yoon, E. (17. October 2021). *Harvard Business Reviw*. Noudettu osoitteesta Demand and Sales Aren't Equivalent: https://hbr.org/2012/10/demand-and-sales-arent-equivalent

Yunishafira, A. (2018). Determining the Appropriate Demand Forecasting Using Time Series Method: Study Case at Garment Industry in Indonesia. 563.

Zhang, Z.;& Moore, J. (2015). *Mathematical and Physical Fundamentals of Climate Change.* Amsterdam: Elsevier.