Erasmus University Rotterdam

Erasmus School of Economics

Master Thesis

Business Analytics & Quantitative Marketing

A predictive extension to the INretail ShoeMonitor

Using Machine Learning to forecast shoe sales

Margot Vos (547814)

Supervisor EUR: Prof. dr. ir. Rommert Dekker

Second Assessor: Dr. Hakan Akyuz

Abstract

Shoe sales forecasting is a challenging task because of the fact the sales are highly fashion trend driven and seasonal. In this study, the weekly sales of shoes for all members of INretail is forecasted by using two models, namely support vector regression (SVR) and LightGBM. SVR has the advantage of reducing the chance on overfitting in combination with promising results. LightGBM is a relatively new and popular algorithm because of the high accuracy and low computation time. The accuracy of both measures is also compared to a naive model. This research shows that LightGBM outperforms SVR in both accuracy and computation time. Moreover, it shows an improvement of approximately 60% with respect to the naive model. Since most historical sales data used is affected by COVID-19, a small case-study is performed to the effects on the forecast. The impact of COVID-19 on the forecast of SVR is negligible, whereas the forecast of LightGBM is more influenced. Removing the affected weeks of data, results in a better forecast for LightGBM.

October 13, 2021

Table of Contents

1 Introduction

INretail is the largest trade organisation for all non-food retail in the Netherlands. They fight for a strong shopping economy on behalf of 13.000 stores. They know the market throughout and they will step in where necessary to make doing business easier. Since its foundation 100 years ago, a lot has changed. Besides growing in the number of members which are connected to INretail, a transition is taking place in terms of strategy. The new strategy focuses on creating a data-hub for all entrepreneurs. Last year, INretail developed a monitor for every retail sector which gives the members insight in the revenue of their company. The first developed monitor is the ShoeMonitor, which focuses on the shoe sector. Therefore, most data is available in this sector and this research will focus on shoe sales. But what tells the past us about the future? Is it possible to make a prediction based on all past data?
Enriching the ShoeMonitor with predictive data is in line with their new strategy. Because all members should be able to access this prediction through an online dashboard which is updated every week, it should not take too long to generate the prediction of the sales.

As regards the current ShoeMonitor, the used benchmark is divided into two categories, namely: the big retailers and the so called MKB'ers, which are the smaller ones with just one or a few stores. The used benchmark is then the average of all entrepreneurs belonging to that specific category. The data is obtained weekly whereby INretail a direct link has with the checkout systems of the stores. For the retailers, it is possible to see its turnover of last week compared to the benchmark and its own turnover a year ago, as well as the cumulative turnover of this year compared to the benchmark and its own cumulative turnover one year ago. Furthermore a division has been made in region, online/ offline sales, customer group and product category. With the great amount of past data which is available via this monitor and multiple forecasting techniques, an accurate sales prediction should be possible to produce. But is it possible to make the forecast accurate enough to use it in other purposes, for example to adjust your procurement process accordingly?

Sales forecasting for the retail sector in general is important but challenging. It is of great importance because accurate forecasts of sales can potentially increase the profit of the retailers by reducing over-stocking and under-stocking costs. Instead of using the so called newsboy model to estimate the costs of over- or under-stocking, an accurate forecast can reduce these costs even further by including trends and seasonality (Lau, 1980). Other interested parties in the supply chain, such as for example the shoe manufacturer, can benefit as well from an accurate predic-

tion ((Sun et al., 2008) and (Ramos et al., 2015)). Despite the importance of sales forecasting, many small and medium retailers in general are not aware of it and use an intuitive approach for the purchase of items in their stores.
Retail sales time series often show strong trend and seasonal patterns presenting challenges in developing effective forecasting models. Moreover, the long time-to-market in contrast to the short cycle of products makes it even harder to come up with an accurate forecast (Singh et al., 2015), but not impossible with today's forecasting techniques. Therefore, a lot of research has been done in the last years in the field of retail sales forecasting. This has resulted in a large number of published forecasting ideas, techniques and measures aiming to produce a prediction as accurate as possible.

In most papers the used techniques are normally based on traditional statistical methods and are either unsatisfactory or inefficient because of having low accuracy and not making use of all external variables (Makridakis et al., 2020a). Whereas, the demand pattern of customers is altered with respect to the holidays, weather, seasonal patterns, economic situations and other variables (Makridakis et al., 2020a). Hence, we will further research which external sources should be added to make the prediction as accurate as possible. Are there other ways of improving accuracy instead of only the addition of more external variables? With the emergence of multiple new algorithms, all with their own specialization, the search continues to find the optimal specification for this task.

Nowadays, more research has been done in the field of machine learning for retail forecasting, whereas in the past forecasting by statistical models was more popular. This is because of the availability of big data. For example, artificial neural networks have been used to predict sales, such as the backpropagation neural network. Although it performs well on accuracy, problems such as over-tuning and long computational times still remain (Sun et al., 2008). In addition, computational times will rise as well when adding the earlier mentioned extra variables. Providing the retailers with an up-to-date forecast dashboard, a model which is not computationally feasible or not accurate is far from desirable. Can we find middle ground in decreasing computation time and increasing accuracy?

What is the effect and usefulness of forecasting when living an unpredictable situation like COVID-19? Such a situation leads to drastic rise of uncertainties in sales and demand. In this case, there is need to have two types of prediction, a short-term prediction and a long term prediction (Khakpour, 2020). Now that the pandemic has been going on for over a year now, some effects on the sales may be visible. These effects will remain visible for a very long time, especially in time series forecasting with lagged variables. Because this virus is still alive, any research into its effects in a specific field is of great value and could be very useful for further research. Thus, besides looking for the best forecasting technique in this situation, another look will be given on how to deal with these affected data.

The described problems above lead to three main research questions which will be answered in this thesis:

'Which machine learning method is best to predict shoe sales?'
'Which external variables should be added to the prediction model to forecast the shoe sales?'
'Which techniques should be used to maintain computational feasibility?'
'What do we know about the effects of COVID-19 on shoe sales and how to deal with them in forecasting?'

The rest of this paper is organized as follows. In the following section, the relevant literature is reviewed. Next, the data will be introduced. After this, the methodology will be explained. After this, the results will be shown. Lastly, a conclusion and a reflection on the research will be given. In addition, some recommendations for further research are provided.

2 Literature

In this section a review will be given on the machine learning techniques already used for sales forecasting. First, an analysis on sales and demand forecasting for the retail sector in general will be given. After this section we will focus on predictive analysis in fashion retailing. To get more clarity about the already known effects of COVID-19 on the retail sector a literature study has been done in the third part of this chapter. In the last subsection, a literature review about the used models in this paper will be given.

In recent years, sales forecasting has been a very popular topic. A lot of research has been done in statistical models, but with the upcoming of big data, machine learning (ML) is on the rise. For many years it has been empirically found that simple methods are as accurate as complex or statistically sophisticated ones (Makridakis et al., 2020b). Researchers preferred statistical methods over ML methods because of the limited data sources, inefficiency of ML algorithms, the need for pre-processing of the data in an ML approach, and restricted computational power. (Makridakis et al., 2018).

The main findings of the M4 and M5 Accuracy competitions, which took place in 2018 and 2020 respectively, indicate that ML methods and hybrid models (where the final model consists of a combination of multiple algorithms) are becoming more accurate than statistical ones and therefore more research should be done in the field of both forecasting options ((Makridakis et al., 2020b) and (Makridakis et al., 2020a)). Keeping in mind the task of this research, creating a model with a relatively short computation time, the choice has been made to just focus on single machine learning models for sales and demand forecasting. The M competitions are large-scale prediction competitions with a large number of participants. All participants get the same prediction task with the same data, therefore the performance of a wide variety of models is studied and the results are very useful for future research. The findings of the latest competition, M5, are used as guideline for this research.

2.1 Sales forecasting for retail industry in general

For this subsection a literature study in the field of sales forecasting in the retail industry in general, such as fast moving consumer goods (FMCG) and electronics, has been done. The focus is on recently published research papers which use different techniques and variables to forecast the sales.

Krishna et al. (2018) conducted a comparison study between a number of machine learning based techniques for food sales prediction of a retail store (Krishna et al., 2018). The data contains information of the food sold by different supermarkets. Moreover, some information about the stores and promotion for the products is given. In this article, various regression techniques are compared against gradient boosting algorithms. The result show that the gradient boosting techniques like AdaBoost and Gradient Tree Boost outperform other linear regression methods (linear regression, polynomial regression, lasso regression, ridge regression) in retail store sales forecasting. The research shows the importance of hyperparameter tuning. Without hyperparameter optimization, the model will not perform as expected.

Kumar et al. (2020) developed a model based on the back-propagation neural network (BPNN) classifier, trained by fuzzy inputs (Kumar et al., 2020). In this research, besides using only historical sales data, numerous variables are added to the model such as a variable based on which advertisements are spread, other expenses, promotions and marketing data. These variables are used to get knowledge about the demand of the product and the effect on the sales. The model is evaluated by comparing it with more common used algorithms (both statistical and machine learning) such as, ARIMA, SVM, and random forest. The study concludes that the fuzzy neural network-based framework has a higher accuracy and provides better forecasting results because of flexible adjustment and capturing linear behaviour of time series

Furthermore that year, Khakpour (2020) applied multiple ML techniques for demand forecasting in the FMCG and retail industry to help demand management and other operative and strategic decisions (Khakpour, 2020). The used ML techniques are: XGBoost, Random Forest, and Support Vector Machine, where XGBoost showed the best outcome for this particular problem. Furthermore, the process of hyperparameter optimization was achieved in a faster way with the help of the Apache Spark as a big data handling tool. However, the research does not tell us something about the size of decrease for the hyperparameter tuning.

Huber and Stuckenschmidt (2020) found that clustering of larger data sets which belong to different time series made it possible to suppress noise (Huber and Stuckenschmidt, 2020). Because of the use of larger data sets, the risk of overfitting becomes less. Clustering was done during the training phase by adding time-invariant features to the time series. So he created a global model which is easier to maintain. This research uses special days on the calendar like holidays as the only external variable, while more variables could be added as well to produce a more accurate forecast.

2.2 Sales forecasting for apparel and fashion industry

A more specific and challenging field of the retail industry as regards forecasting is the apparel and fashion industry. The sales of those products are highly seasonal (for example tank-tops which are mostly sold during summer and sweaters are mostly sold in winter) and fashion driven. Moreover, the products of these industry are essential but all with another lifespan, for example sneakers worn daily often have a shorter life cycle than smart shoes for special occasions. Those aspects makes the forecasting task more complex. The past years, some research has been done to find a forecast model which is as accurate as possible for those specific patterns in these time series. Sophisticated techniques such as extreme learning machine (ELM) have enabled researchers to extract more features from the historical data, even if the data contains a lot of noise. In the paper of Thomassey (2013) a lot of research has been done in the field of external variables (Thomassey, 2013). He concludes that the impact of each of these variables is especially difficult to estimate and the impact of each of these variables is not constant over time. Hence, more research in the field of external variables should be done. Furthermore, the demand pattern for fashion items is influenced by a lot of factors. Unfortunately, it is difficult to capture these effects. Thus, we will have to take into account that a fairly accurate forecast can be made for this task with an bigger error interval than than when the same methods are used for other forecasting purposes. Sun et al. (2008) applied ELM as well, but with an extension (Sun et al., 2008). The known advantages of ELM in comparison with more traditional gradient based algorithms are the high generalization performance; avoidance overfitting; no need for determination of stopping criteria, learning rate, learning epochs and local minima. The research shows that forecasting sales in fashion retailing by using both ELM and it extension can produce smaller prediction errors than some other well performing forecasting methods in the literature, namely two versions of backpropogation neural networks (BPNN).

Ramos et al. conducted a research in the field of shoe forecasting (Ramos et al., 2015). They performed a forecast with different exponential smoothing methods as well as ARIMA models. They compared the outcomes of all models when applied to data of retail sales of five different categories of women footwear from a Portuguese retailer. To evaluate their results of the error, trend and seasonality (ETS) and ARIMA models they used three different performance measures, namely RMSE, MAE and MAPE. They concluded that the results for a one-step ahead and multi-step ahead based on these performance measures do not differ much. For both the ETS and ARIMA model the multi-step ahead forecasts are generally better than the one-step forecasts because it incorporates information that is more updated. Therefore the preference is with a multi-step forecast to give the retailers a forecast which is as accurate as possible. For this specific task, only statistical models are taken into account, while other researches state that machine learning methods outperform statistical models. Therefore , research in this specific field using machine learning should be conducted and the stated conclusions should be reviewed.

2.3 The effect on retail sales of COVID-19

At 27th of February, patient zero for COVID-19 was detected in the Netherlands. In March 2020 the virus was confirmed to be a pandemic. The following months measures were taken to prevent the spread of the virus. Both the images of overcrowded hospitals and high death rates due to the virus as well as the new measures resulted in a change of the shopping behaviour of people. The gap of knowledge raised a lot of uncertainty. People got scared to be in crowded places like shopping areas. Eventually all retailers were forced to close their stores on the 13th of December 2020. In March 2021 customers could come the the shops again with an appointment and on the 27th of April, you could shop again without an appointment but with some restrictions.

Sales forecasting in times of a pandemic is very difficult because of the uncertainty of the spread of the virus and the duration of the virus. People nowadays have never been in a pandemic of this size before and the limited availability of historic data makes it hard to make an accurate forecast.

Most research done in the field of COVID-19 is about outbreak prediction. Different researchers such as epidemiologists, mathematicians, decision scientists and operational researchers have applied multiple models (both time series and machine learning models) for predicting the spread of the virus (Nikolopoulos et al., 2021). These distribution predictions can help predict other time series, such as retail sales. However, much is also determined by the government. They can, for example, choose to unexpectedly close the stores or take other measures that have a major impact on retailers and their sales. This makes forecasting even more difficult. The closing of stores and other measures lead to different shopping behaviour. Some retailers have liked click & collect (purchase cloths online and collect them in store) so much that they want to continue with it despite the reopening of the stores (source: INretail member poll March 2021). Therefore the effect of COVID-19 on checkout sales will be long-lasting.

Another research field found in literature already is the effect of the pandemic on the supply chain (Ivanov, 2020). It gives insight in the prediction of recovery time of the supply chain after an epidemic outbreak as well as the impact on the supply chain, both short-time and long time. The effect and how to deal with supply chain disruptions with other causes has been researched

by (Wu et al., 2007). They presented a 'Disruption Analysis Network'. Using this model, better management of the supply chain during the disruption will be accomplished which lead to quicker responses to the customers, lower missed costs for multiple parties due to for example lower inventories and a reduced bullwhip effect for both retailer and manufacturer.
However, predicting the direct and indirect effect on sales is still to be investigated.

2.4 SVR and LightGBM

With the great amount of available forecasting techniques nowadays, the search for the best performing algorithm for this specific task gets more difficult. After reviewing the literature, the choice has been made to perform the shoe sales forecast using two methods, namely support vector regression (SVR) and LightGBM. Where lightGBM is a relatively new developed algorithm and SVR is a well-known and well-performing algorithm for forecasting ((Ke et al., 2017) and (Wen et al., 2014)). Due to the high risk of overfitting when predicting with regression models based on sample data, SVR became more popular (Yu et al., 2013). SVR searches for the optimal configuration to reduce the chance on overfitting and achieve high model prediction accuracy. Yu et al. (2013) used SVR for forecasting newspaper and magazine sales. Those sales are today highly influenced by the availability of news on the internet (Yu et al., 2013). They recommend further research in the selection of explanatory variables. Yang et al. (2007) also used SVR for sales forecasting (Yangl et al., 2007). The research was conducted for the Chinese tobacco industry and showed promising results, especially in efficiency by using less CPU and memory. Because of the big amount of retailers for which a forecast must be made, less CPU and memory usage is desirable.
As regards the best performing models in the M5 forecasting competition, most of them utilized LightGBM (Makridakis et al., 2020a). In the competition, sales data of Walmart has been used to forecast daily sales. In this research the performance of LightGBM on the highly seasonal and fashion driven shoe sales will be tested. Concerning the results of the competition, improvement can be made by gaining knowledge on how the ML methods create their prediction. This is important when for example managers really want to use it, because they are generally unwilling to make decisions on this forecast when they do not understand how this prediction is produced. In addition, further research in the field of extra explanatory variables such as price, promotions and special events is also desirable.
As said before the main task is to develop a model which is both accurate and computationally feasible. The possibility of achieving this combination for shoe sales forecasting will be researched using SVR and LightGBM.

To conclude, although the overall good performance by multiple sales forecasting models in the papers mentioned in this literature section, future research could be done in the field of hyperparameter optimization and addition of external variables such as holidays. Those aspects will be added to the overall well-functioning forecasting models SVR and LightGBM to produce a shoe sales prediction for all shoe retail members of INretail. The accuracy of both models will be tested to form an idea whether forecasting shoe sales is currently possible for INretail.

3 Data

In this section, the by INretail provided data will be introduced. Thereafter an exploratory data analysis has been done to get some insights in the past data. First, we will have a look at the total sales per shoe type, after this we will have a look into all given variables separately in the given data set. Moreover, fieldwork has been done to get some feeling of the sector. A short summary is given in section 3.5. Lastly, information about the performance criteria will be given.

3.1 Provided data

Nowadays most stores in the Netherlands have electronic checkout systems. Via a direct link, INretail gets weekly checkout data of 526 shoe stores. These shops belong to 74 different entrepreneurs. This data is available from the first week of January 2018 until June 2021. A lot of data is affected by COVID-19. As mentioned before in the literature section, COVID-19 started in the Netherlands at the 27th of February 2020. Although the shops stayed open until the 15th of December 2020, people got reluctant to go to crowded places like shopping areas.

The data set consists of information about the sales in euros, shoe type, category, number of receipts (given per week for each brand-shoe type combination), number of lines on the receipt, brand, the four numbers of the ZIP-code of the store, the suggested retail price and in 2018 also the purchase price of the shoes for the retailers. For readers convenience, all mentioned sales will be in euros. As a result of the given suggested retail price and total sales realized, the discount rate could be calculated per week for each brand-shoe type combination.

The realized sales can vary a lot. For the big retailers, weekly sales of a certain shoe type can reach more than €2000000,-, while for the small retailers a weekly sales of a few tens is reached. When we take a look at a specific shoe type summed over all retailers, the weekly realized sales can almost reach €9000000,-, while other shoe types are only sold once a week. Moreover the data contains some lines with negative realized sales. This means that a pair of shoes is returned in the store and the amount is refunded to the customer.

To use this data for forecasting, data cleaning must be done beforehand.

3.2 External data

Besides the data provided by INretail, external data sources will be utilized. The planning of the bank holidays in the Netherlands will be added to the model, as well as the Dutch school holidays by using the `Python` package `workalender.europe`. Lastly, the predefined astronomic seasons of the Northern Hemisphere will also be added to the model. Therefore, the dates as provided by National Geographic Encyclopedia are converted to ISO-weeks (source: National Geographic Encyclopedia).

3.3 Exploratory Data Analysis

In this study we analyze the time series of the sales of 74 retailers divided in 24 shoe types and 7 categories (men, women, children, boys, girls, other and not registered). Not all retailers sell all shoe types for each category. When we take a look at each combination of category and shoe type, we get 168 combinations. If we take a look at the shoe sales per shoe type for a retailer, we have 1944 time series.

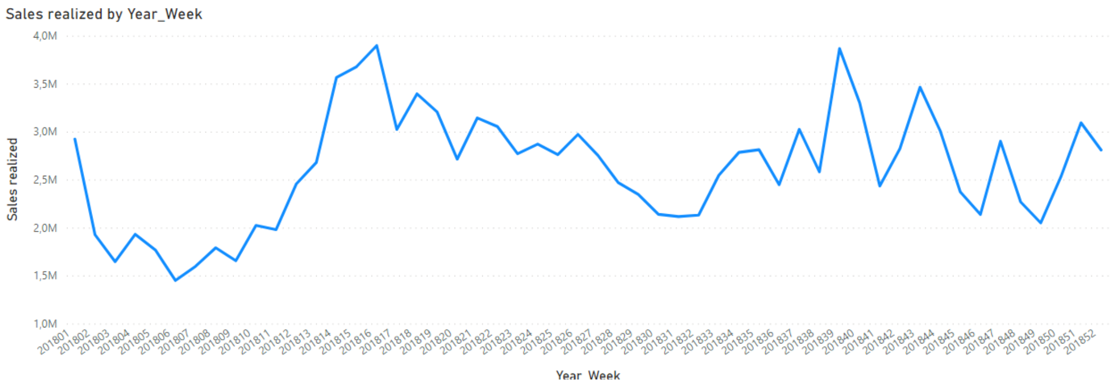


Figure 1: Plot of total realized sales per week for 2018

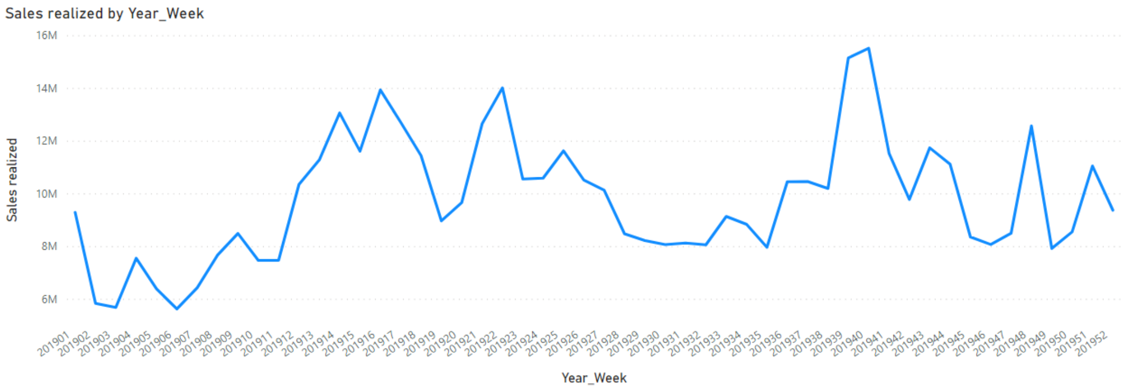


Figure 2: Plot of total realized sales per week for 2019

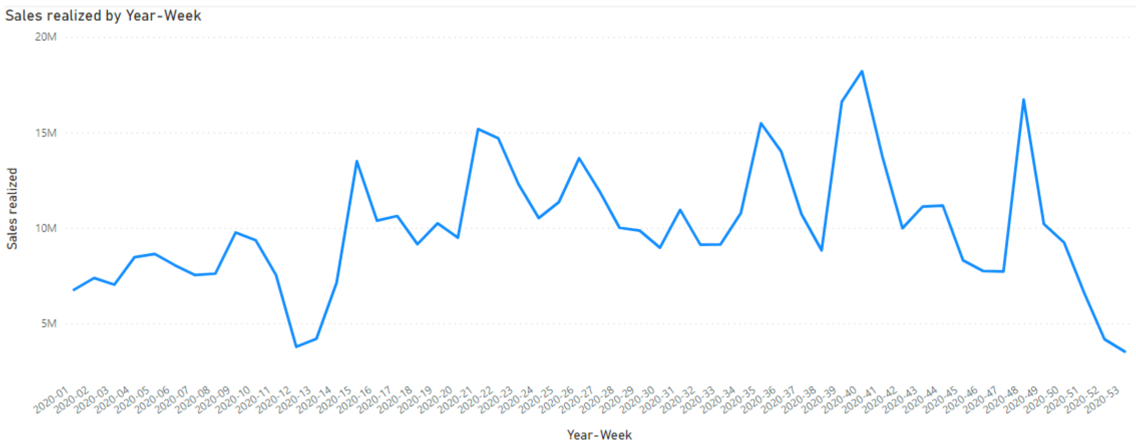


Figure 3: Plot of total realized sales per week for 2020

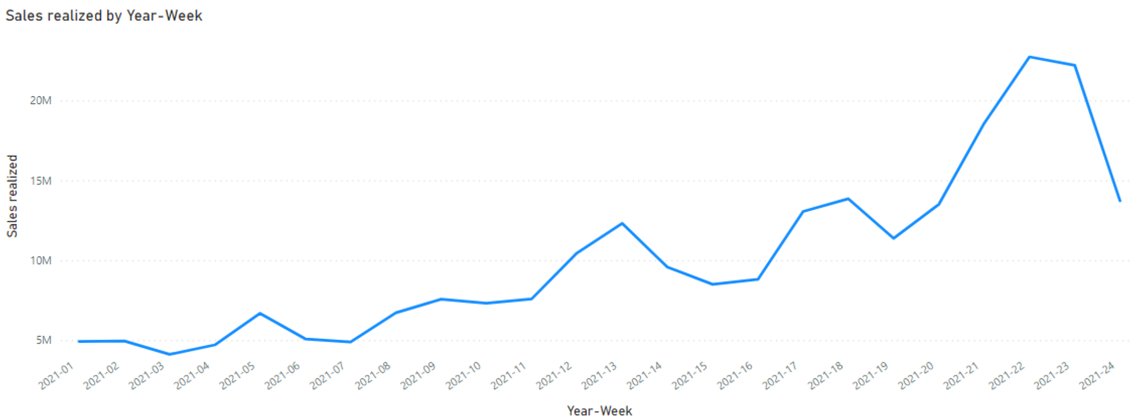


Figure 4: Plot of total realized sales per week for 2021

First, let's take a look at the time series of the realized sales plots of all years separately in figures 1, 2, 3 and 4. It is clear to see that the time series of 2018 and 2019 follow the same pattern. Both series have clear outliers at week 16, 39, 43, and 51. In the time series of 2019 an extra outlier can be found in week 22. The total sales in 2018 are much lower because the big retailers like 'Van Haren' and 'Van den Assem' began to share their data since 2019. In 2020 a great fall in sales can be seen in week 12 and week 13. This is during the beginning of COVID-19 in the Netherlands. Another great fall of sales can be found in the last weeks of 2020. In these weeks, COVID-19 peaked and all retailers had to close their stores. In the first weeks of 2021 a slow recovery of the sales can be found, because from week 8 click and collect was possible and in week 14 all entrepreneurs were allowed to open their stores again, albeit in modified form. More information about the handling of data which is polluted by COVID-19 can be found in section 3.4.

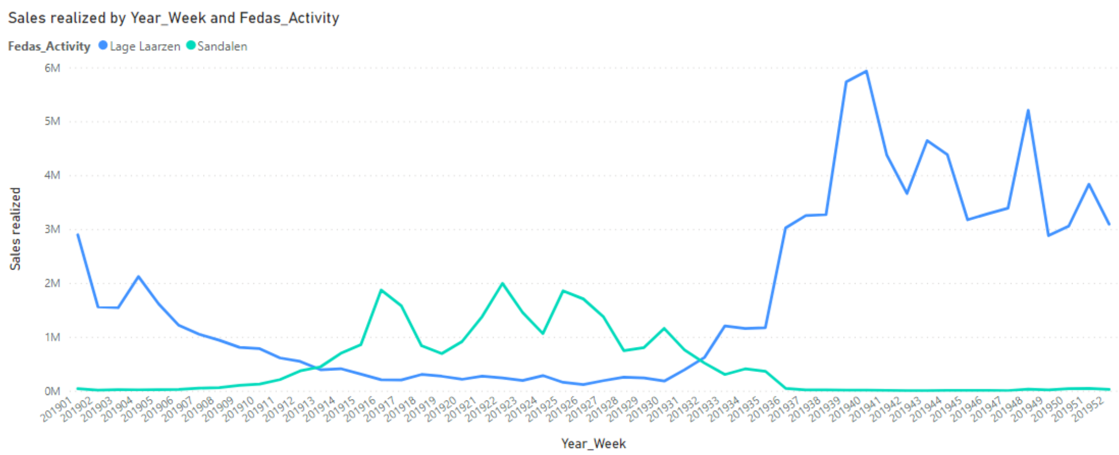


Figure 5: Plot of total realized sales per week for 2019 for booties and sandals

To gain a better understanding of certain variables, we now take a closer look at the year 2019. This year contains most data and is not affected by COVID-19. The time series are now plotted per product type. As an example in figure 5 the sales of 'booties' and 'sandals' are plotted. The difference between seasons of these types is clearly visible in this plot. Besides this, it becomes clear that if we take a look at the outliers of the booties category in figure 5, they coincide with the outliers in figure 5. The 'flats', 'espadrilles/ loafers', 'sandals', 'flip flops' and 'pumps' are mostly sold during the summer, while boots, booties and slippers are mostly sold during the winter. For the other categories, no great change in sales per season can be found. We say that the summer starts in week twelve and ends in week 37, the winter starts in week 38 and ends in week 11 (source: National Geographic Encyclopedia). The series mentioned above present strong seasonality and all series are obviously non-stationary.

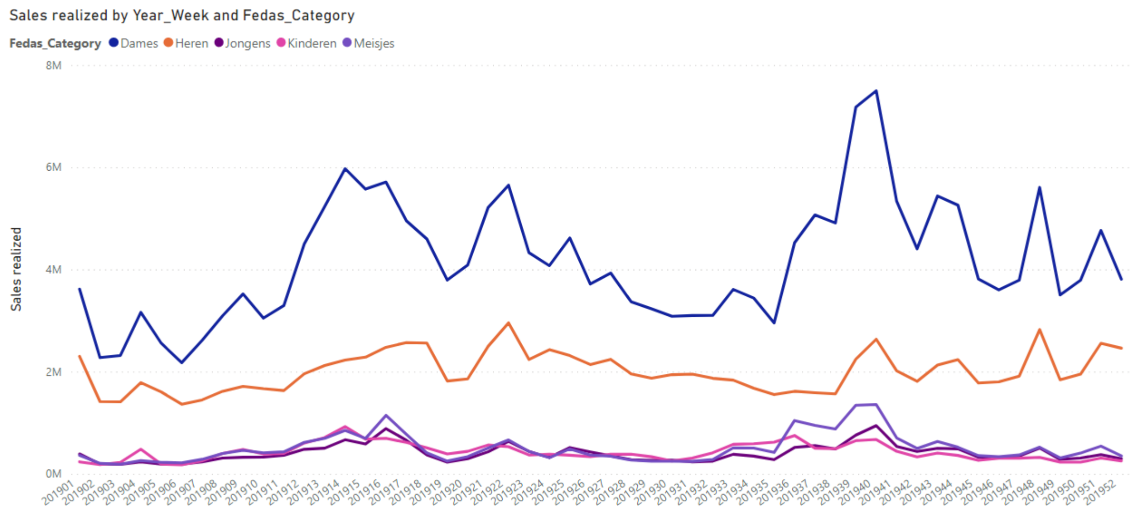


Figure 6: Plot of total realized sales per category for 2019

If we divide the data into categories (figure 6), we see that the overall sales for 'women' is much higher then all other categories. In addition, no big difference can be found in the pattern and sales of the categories 'children', 'boys', 'girls'. Therefore, these data will be merged into one category: children.

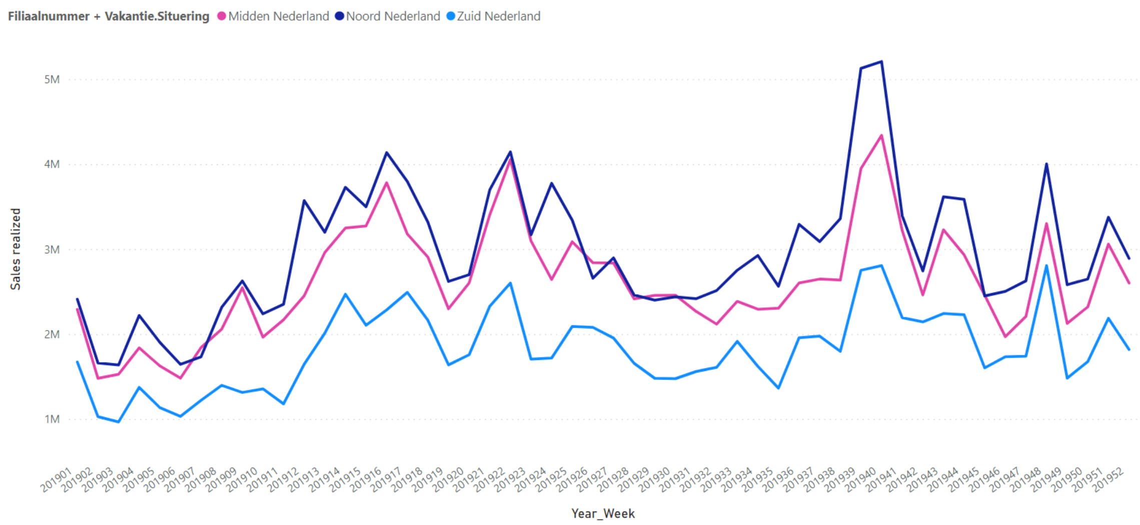


Figure 7: Plot of the total sales separated by different holiday regions

In the Netherlands, the start of the school holidays depends on your location. We are dealing with three different regions: north, middle, south. A plot has been made with a distinction per region for 2019. All regions follow almost the same pattern, except for some weeks. For example, the north and middle region of the Netherlands (dark blue and magenta respectively) show a peak at week 16 of 2019, while for the south of the Netherlands (light blue) shows a peak at week 17 of 2019. Furthermore, a lower total sales occurs during the summer period.

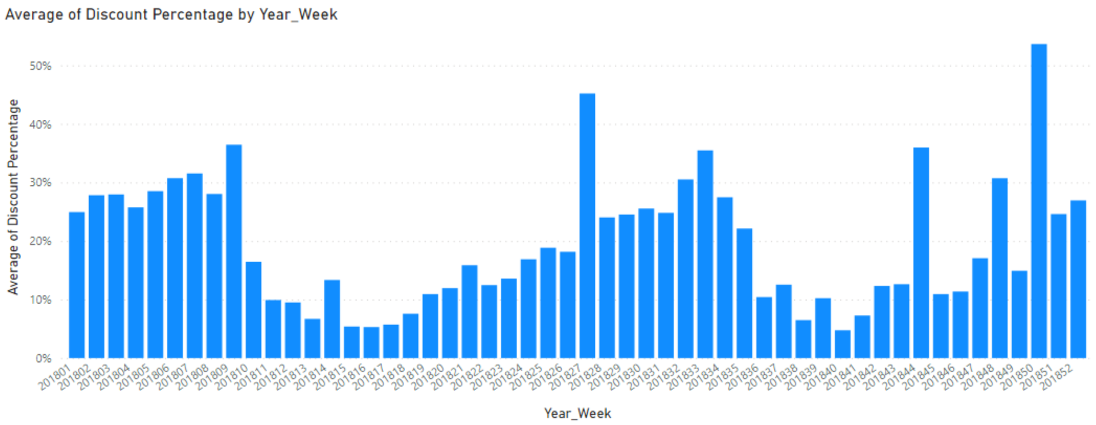


Figure 8: Histogram of discount rates per week in 2018

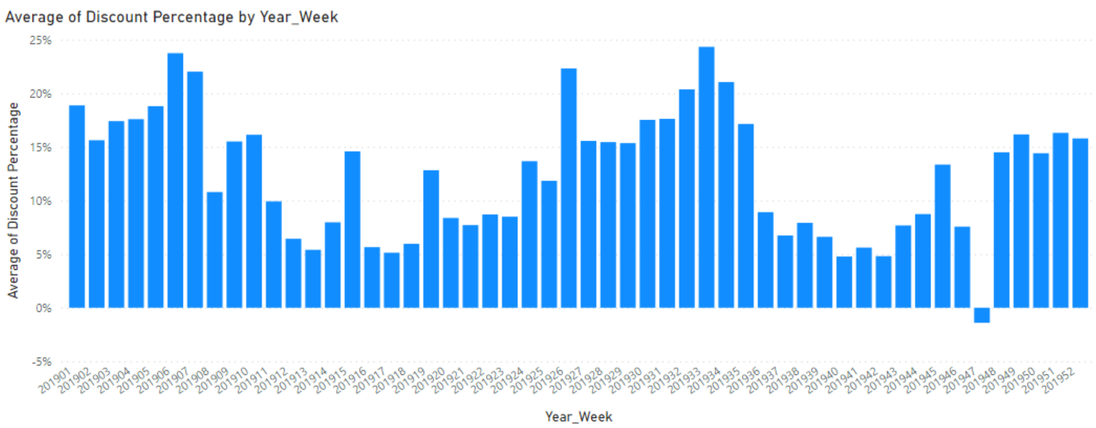


Figure 9: Histogram of discount rates per week in 2019

The last plots which have been made for the exploratory data analysis are two histograms with the average discount rate per week found in figures 8 and 9. Again for 2018 and 2019 the same pattern could be found for the two years. We distinguish two main peaks: one at the beginning of the new year which already starts in the last weeks of past year and one peak at approximately week 34. This can be related to the sales, at times of higher discount the sales are relatively low. In the discount plot of 2019 a negative value can be found for week 47 in 2019. This means that in this specific week the average discount rate was negative, thus the realized sales were higher than the given suggested retail price.

For the given data the discounts percentages will be grouped in three classes found in table 1.

Table 1: All possible discount classes used as dummy variables

	Class
10 % - 30 %	Low discount
30 % to 50%	Medium discount
50% to 75%	High discount

These classes will be transformed to dummy variables and taken into account as discount lags for the prediction.

The findings of this exploratory data analysis will be taken into account for the prediction of shoe sales, such as a variable which include the spread of holidays for different regions, as well as a variable with a lag of one year and a lagged dummy variable for the discount.

3.4 COVID-19

Some effects of the virus on past data are already mentioned in section 3.3 Exploratory Data Analysis. This makes forecasting on basis of this data challenging. The literature review showed that not much research has been done on the effects of COVID-19 on retail sales. Therefore the choice has been made to take a deep-dive into the data affected by COVID-19. First the data will be used without taking into account the effects of the pandemic. Thus, the data will be used without adjustments. Next, the weeks in which a great effect of the pandemic can be seen in comparison with other years, are removed and interpolation is done. Both methods will be compared to give some insights in the effect of COVID-19. In figure 10 a plot can be found with the total sales of year 2019, 2020 and 2021 partly. With some exceptions the years follow the same patterns. For these exceptions, it was investigated whether COVID-19 could have played a role in this, for example by the announcement of new measures. For these weeks a data adjustment will be done to measure the primary effects of the pandemic on forecasting. The weeks where a clear disruption is visible in the data will be removed and the other weeks will be interpolated. This means that weeks 12 and 13 of 2020 will be removed as well as week 50 in 2020 till the ninth week of 2021.

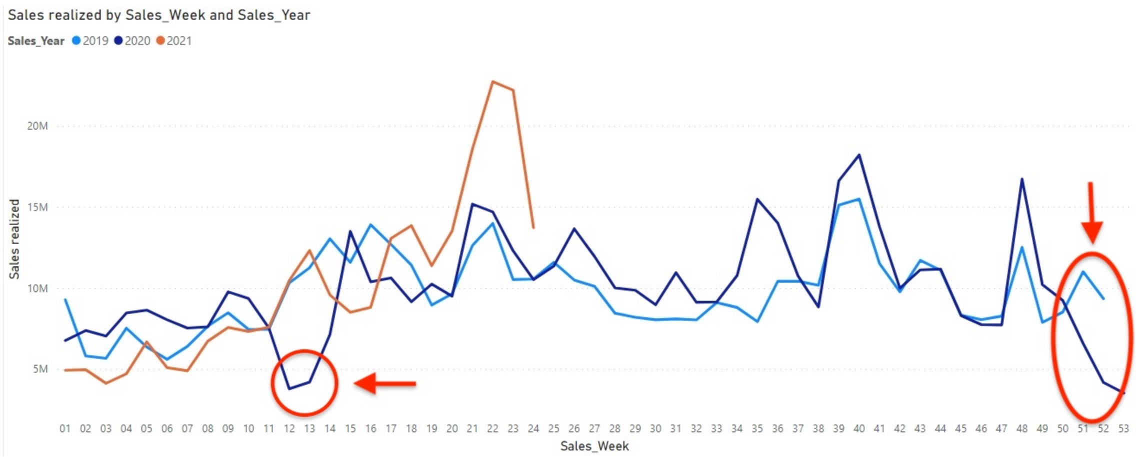


Figure 10: Plot of total realized sales for 2019, 2020 and 2021 partly. The disrupted weeks are highlighted by red circles

3.5 Fieldwork shoe retailers

To get more grip on the buying and selling of shoes, fieldwork has been done. This gives insight in the procurement process of the shoe retailers as well as the wishes for the new extension of the 'ShoeMonitor'. Shoe retailers mostly base their procurement on the sales of last year. 80% is bought almost a year in advance. For example, the collection of spring/summer 2022 is mostly purchased August/September 2021. These months the retailers sold most of this years collection and have a good insight in the which shoes were popular and which shoes did not sell as planned. The other 20% can be purchased at the manufacturer during the season. Many retailers lose grip on their procurement and trust completely on the experience of big manufacturers. For these retailers the predictive dashboard will help to control their purchases themselves.

3.6 Performance criteria

To evaluate the accuracy of the forecast, a naive based model will be used. This means that the forecast of this week is the same as the sales of the exact same week last year. Nowadays most entrepreneurs now base their procurement on this, therefore this model will be used as a benchmark.

To prevent overfitting, the accuracy of the forecast will be determined on data which has not been used when fitting the data. Therefore, the data will be split up in two parts: one part for training the data and the part for testing. For each time series the same train-test split has been used. The data has been split in a ratio of 70% for training and 30% for testing. This means that 89 weeks will be used for training and 39 weeks will be used for testing. The train data will be both used for model fitting as well as the selection of hyperparameters. The test data will solely be used to verify the fit of the model.

4 Methodology

4.1 Introduction

The main tasks for this research are improving forecast accuracy with respect to the naive model and develop a model with a feasible computation time to use it in the 'ShoeMonitor'. It should not be the case that every week the model needs a few days to produce the forecast. To accomplish this main tasks, two models will be used. The fist model is the longer existing Support Vector Regression. After this, the relatively new gradient boosting algorithm 'LightGBM' will be introduced. Lastly, the measures for evaluation of the accuracy will be given.

4.2 Support Vector Regression

The first described model, is the support vector regression (SVR). This following description is based on the method of Alex Smola and Bernhard Schölkopf (Smola and Schölkopf, 2004).
With the emergence of big data, a lot of research has been done in SVR. A support vector regression is highly preferred by many as it produces significant accuracy with less computation power. SVR has been successfully used for machine learning with large and high dimensional data sets ((Wu, 2009) and (Pillo et al., 2016)). SVR is a combination of support vector machines (SVM) and regression functions. SVMs are used for classification problems. The idea of SVM is to find an acceptable error range, instead of finding the exact error term. The basic idea of SVM for function approximation, in other words SVR, is mapping the data into a high-dimensional feature space by nonlinear mapping and then performing a normal linear regression in the feature space. The non-linear SVR model is:

$$f(x) = w^T \phi(x) + b \tag{1}$$

Where b is the deviation vector, w is weight and $\phi(x)$ is a nonlinear mapping function. In order to ensure the optimization problem, slack variables θ and θ^* will be used:

$$\phi(w, \theta, \theta^*) = \frac{1}{2}||w||^2 + C(\sum_{i=1}^{i} \theta_i + \sum_{i=1}^{i} \theta_i^*)$$

$$\text{s.t.} \begin{cases} y_i - w^T \phi(x_i) - b \leq \epsilon + \theta_i, & i = 1, 2, ..., N \\ w^T \phi(x_i) + b - y_i \leq \epsilon + \theta_i^*, & i = 1, 2, ..., N \\ \theta_i \geq 0; \quad \theta_i^* \geq 0 \end{cases} \tag{2}$$

Where C indicates the regularization coefficient which controls the errors and ϵ indicates the regression. With the aim of solving the SVR, the objective function should be combined to Lagrange function constraints to make the nonlinear problem into a dual program:

$$max[-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(a_i-a_i^*)K(x_i,x_j)+\sum_{i=1}^{n}a_i(y_i-\epsilon)-\sum_{i=1}^{n}a_i^*(y_i-\epsilon)] \quad \text{s.t} \quad \begin{cases} \sum_{i=1}^{n}(a_i-a_i^*)=0 \\ 0\leq a_i\leq C, 0\leq a_i^*\leq C \end{cases} \tag{3}$$

where a_i and a_i^* are the so-called Lagrange multipliers. The K indicates a kernel function. After introducing the Lagrange multipliers and exploiting the optimality constraints, the function form equation 1 becomes:

$$f(x)=\sum_{i=1}^{n}(a_i^*-a_i)K(x_i,x)+b \tag{4}$$

Multiple kernel functions exist, but the most commonly used in literature is the radial basis function (RBF):

$$K(x)=exp(-||x-x_i||^2/2\sigma^2) \tag{5}$$

For this study, RBF will be used as well.

Grid search and cross validation

In this paragraph the grid search to find the optimal specification will be discussed. To find the optimal specification for each time series, the parameters need to be tuned. The grid search will take each possible combination of parameters and calculates the R^2. Eventually, the model chooses the specification with the lowest average R^2 for the final prediction. The list of parameters and its options can be found in table 2

Table 2: Parameter value options used in grid search for optimal model specification.

	Possible values
C	1e6 , 1e7, 1e8, 1e9
ϵ	1e-5 , 1e-4, 1e-3, 1e-2, 1e-1

The parameters ϵ and C and the kernel parameter σ can also be found in equations 4.2 and 5 respectively. If ϵ is given, increasing C has the effect of making more often the correct output for the function realized by the SVR: however, if exaggerated, the model is probably overfitted and will therefore fit the training data very well, but will probably not have a high accuracy as regards the test data (Pillo et al., 2016). The SVR will lose the generalization properties with respect to samples not in the training set. Despite the grid search, overfitting the data is still possible. To reduce the chance of overfitting any further k-fold cross validation will be added with $k = 5$.

4.3 Light GBM

Gradient-boosted regression trees have gained much popularity in recent years. A relatively new algorithm is LightGBM (LGBM) (Ke et al., 2017). The following description of LGBM is based on the method by Ke et al. (2017). Like any boosting algorithm, it trains a series of simple models $f_k(x)$ (i.e., decision trees) based on the obtained residuals of the previous model $L^{(t)}(y, \hat{y}^{(t-1)} + f_t(x))$. Hence, the prediction is the sum of all the trained simple models. The advantage of LGBM is that it does not grow a tree level-wise as most other implementations do, instead it grows trees leaf-wise. It chooses the leaf which is assumed to yield the largest decrease in loss. The LGBM implements a highly optimized histogram-based decision tree learning algorithm, which yields great advantages on both efficiency and storage of unused branches, instead of decision trees which search for the best split point between different sorted features. Therefore, LGBM is suitable for larger datasets. The LGBM algorithm utilizes two novel techniques called Gradient-Based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) which allow the algorithm to run faster while maintaining a high level of accuracy.

With GOSS, only the greater gradients will be used to estimate the information gain and the relatively small instanced will be removed from the model. It is proven by Ke et al. (2017) that 'since the data instances with larger gradients play a more important role in the computation of information gain, GOSS can obtain quite accurate estimation of the information gain with a much smaller data size' (Ke et al., 2017). With EFB, features which hardly take nonzero values at the same times (sparse features) will be bundled, to reduce the number of features. It is so to say an automatic feature selector. A greedy algorithm has to be used to achieve a good approximation ratio (and thus can effectively reduce the number of features without hurting the accuracy of split point determination by much). Ke et al. (2017) found that 'LightGBM speeds up the training process of conventional GBDT by up to over 20 times while achieving almost the same accuracy' (Ke et al., 2017).

In a more mathematical way, LGBM aims to find an approximation $\hat{f}(x)$ to a certain function $\overset{*}{f}(x)$ which minimizes the expected value of a specific loss function $L(y, f(x))$ as follows:

$$\hat{f} = argmin_f E_{y,X} L(y, f(x)) \tag{6}$$

Thereafter, LGBM integrates a number T regression trees $\sum_{t=1}^{T} f_t(X)$ to approximate the final model, which is:

$$f_T(X) = \sum_{t=1}^{T} f_t(X) \tag{7}$$

The regression trees could be expressed as $w_{q(x)}, q \in \{1, 2, ..., J\}$, where J denotes the number of leaves, q indicates the decision rules of the tree and w is a vector which expresses the sample weight of leaf nodes. Hence, LGBM would be trained in an additive form at step t as follows:

$$\Gamma_t = \sum_{i=1}^{n} L(y_i, F_{t-1}(x_I) + f_t(x_i)) \tag{8}$$

In LGBM, the objective function is rapidly approximated with Newton's method. After removing the constant term in equation 8 for simplicity, the formulation can be transformed as follows:

$$\Gamma_t \cong \sum_{i=1}^{n} (g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)) \tag{9}$$

Where g_i and h_i denote the first- and second-order gradient statistics of the loss function. Let I_j denote the sample set of leaf j, and equation 9 could be transformed as follows:

$$\Gamma_t = \sum_{j=1}^{j} ((\sum_{i \in I_j} g_i) w_j + \frac{1}{2} (\sum_{i \in I_j} h_i + \lambda) w_j^2)) \tag{10}$$

For a certain tree structure $q(x)$, the optimal leaf weight scores of each leaf node $\overset{*}{w}_j$ and the extreme value of Γ_K could be solved as follows:

$$\overset{*}{w}_j = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{11}$$

$$\overset{*}{\Gamma}_T = -\frac{1}{2} \sum_{j=1}^{J} \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} \tag{12}$$

Where $\overset{*}{\Gamma}_T$ could be viewed as scoring function that measures the quality of the tree structure q. Finally the objective function after adding the split is:

$$G = \frac{1}{2} (\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I_I} g_i)^2}{\sum_{i \in I_I} h_i + \lambda}) \tag{13}$$

Where I_L and I_R are the sample sets of the left and right branches, respectively.

In conclusion, the main advantages of LGBM against other ML techniques for forecasting found by the M5 forecasting competition are: 'it allows the effective handling of multiple features (e.g., past sales and exogenous/explanatory variables) of various types (numeric, binary, and categorical), is fast to compute compared to typical gradient boosting (GBM) implementations, does not depend on data pre-processing and transformations, and requires the optimization of only a relatively small number of parameters (e.g., learning rate, number of iterations, maximum number of bins that feature values will be bucketed in, number of estimators, and loss function)' (Makridakis et al., 2020a). These advantages are highly vulnerable for the provided data set and the purposes of this research.

Grid search and cross validation

LightGBM contains a multiple tunable hyperparameters. Like SVR, again a grid search will be executed to find the optimal specification for each time series individually. Again, RMSE will be used for scoring and thus the final specification will be based on the specification which has the lowest RMSE for that specific time series.

Table 3: Parameter value options used in grid search for optimal model specification.

	Possible values
Maximum number of leaves	`5, 10, 15, 20`
Minimal data points in leaf	`5, 10, 20, 30, 50, 100, 200`
L1 regularization term on weights	`0, 0.5, 1, 1.5, 2`
L2 regularization term on weights	`0, 0.5, 1, 1.5, 2`

The number of leaves controls the complexity of the tree model. Unconstrained maximum number of leaves can induce over-fitting. Over-fitting can also occur when the minimal data point in leaf is not specified. This depends on what to forecast. A more general forecast (for example for shoe types) will contain more data points and therefore the minimal data points in a leaf must be higher than for a more specific prediction (for example a prediction on entrepreneurial basis).

Tuning is done on both the shoe type level, which is used to make a prediction for each shoe type, as well as on entrepreneurial level, which is used to make a prediction for each retailer individually per shoe type. Therefore, a wide range in hyperparameters is chosen. Every week when new data comes in, the grid search will be executed again.

4.4 Performance measures

To evaluate the forecast, five performance measures will be used, namely the root mean squared error (RMSE), the mean absolute error (MAE), the mean absolute error percentage (MAEP), the R^2 and adjusted R^2. The test data will be used for evaluation of the model. The weekly predicted sales will be compared with the actual weekly sales of each time series from the test data to assess the accuracy of the models per time series.

The RMSE measures the standard deviation of the errors, whereas the MAE the absolute difference measures between the real and predicted data. The RMSE and MAE will be compared to the RMSE and MAE of the naive model. This will be expressed in percentual changes because the size of sales could differ a lot for each shoe type and retailer (see section 3.1) which makes the difference in RMSE and MAE meaningless. This gives the retailers an approximation of the profit which could be made by using this forecast instead of the used kind of naive model in which their procurement is based on the sales of the year before. The motive behind the choice of both RMSE and MAE is the fact that the RMSE is highly sensitive for aberrant observations. The MAE does not have a squared term so the model performance suffers less from outliers. The RMSE and MAE can be written as:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{T}} \tag{14}$$

and

$$\text{MAE} = \frac{\sum_{t=1}^{T}|y_t - \hat{y}_t|}{T}, \tag{15}$$

where in both measures $\hat{y}$ is the forecasted sales and y is the actual sales for a predetermined period which overlaps the period of the test data.
The naive model has the following specification:

$$y_t = y_{t-52} \tag{16}$$

To calculate the percentual changes with respect to the naive model, the following equations will be used:

$$\text{Percentual Change RMSE} = \frac{\text{naive RMSE} - \text{model RMSE}}{\text{naive RMSE}} * 100 \tag{17}$$

and

$$\text{Percentual Change MAE} = \frac{\text{naive MAE} - \text{model MAE}}{\text{naive MAE}} * 100, \tag{18}$$

where 'model RMSE' and 'model MAE' are the obtained RMSE and MAE from SVR and LGBM models. A positive percentage means an improvement with respect to the naive model.

In addition, the MAE is compared with the mean sales of the test data ($\boldsymbol{\bar{y}_{test}}$). This gives understanding in the order of magnitude of the MAE. This measure is slightly different to the mean absolute percentage error (MAPE), in which the MAE is divided by the real sales each week and the MAPE is then the mean of those values, because in some weeks zero sales can occur. Therefore, we have chosen to create the MAEP. For the MAEP, the following equation has been used:

$$\text{MAEP} = (\frac{\text{MAE}}{\bar{y}_{test}}) * 100 \tag{19}$$

R^2 is also called the coefficient of determination and it represents the fraction of variance in

the dependent variable which is explained by the model. Thus it gives knowledge in the general fit of the prediction in comparison with the real values. The R^2 is a value smaller than 1 and independent of the size of the sales. A value for the $R^2_{(adj)}$ smaller than 0 indicates a very bad fit. If for example the R^2 of a model is 0.50, then approximately half of the observed variation can be explained by the model's inputs.
The adjusted R^2 modifies the R^2 for the number of independent variables in the model. Therefore it will always be less than or equal to the R^2. The value of adjusted R^2 decreases if the increase in the R^2 by adding a variable is not significant. R^2 and adjusted R^2 can be written as:

$$R^2 = 1 - \frac{\sum_{t=1}^{T} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{T} (y_t - \overline{y}_t)^2} \tag{20}$$

and

$$R^2_{adj} = 1 - (\frac{(1 - R^2)(t - 1)}{(t - k - 1)}), \tag{21}$$

again $\hat{y}$ is the forecasted sales and y is the actual sales for a predetermined period which overlaps the period of the test data. Moreover $\overline{y}$ is the mean value of the real sales of the test data and k is the number of independent variables in the data.
To conclude, a lower value of RMSE, MAE and MAEP is desirable, in contrast to R^2 and R^2_{adj} where a higher number implies higher accuracy. The RMSE and MAE are to compare the LGBM and SVR with the naive model and the MAEP, R^2 and R^2_{adj} say something about the general fit and explain how well the independent variables of the model explains the variability in the sales. By using all five measures, something can be said about both the possible profit improvement as well as the more general fit of the model.

5 Results

The results chapter is subdivided in different sections. First a comparison will be made about the performance of the SVR model and LightGBM. A distinction will be made between a more broad approach on shoe type basis and the prediction for each entrepreneur individually subjected to shoe types as well. This approach has been chosen because the size of the realized sales on entrepreneurial level is significantly smaller for some retailers and can contain more fluctuations. For example, if only a few shoes are sold during a season of a specific shoe type in a small shop, the data is more noisy. This will make prediction of these time series harder. These fluctuations are more stabilized if we take a look at the overall sales of shoe type.
After this, we will dive deeper into the results of each model separately. Again, a distinction will be made between the forecast for each shoe type and the prediction for each retailer subjected

to shoe type. For each model and configuration (shoe types in general and focused on each retailer individually) a few time series will be highlighted which gives us understanding about the performance when 1.) a clear trend is visible 2.) the data contains outlier(s) and 3.) the time series indicates strong seasonality. Moreover, the time series with the best fit will also be shown. Lastly, we will take a look at the influence of COVID-19 on both models separately.

5.1 Comparison SVR and LightGBM

In this section, a comparison between the performance of the SVR and LightGBM will be given. Both models will be compared to the naive model in which the prediction of a specific week is solely based on the sales of the same week a year before. For this, the test-data will be used as specified in section 4.4. In addition, the MAEP (see section 4.4) will be given as well for both models.

First the sales of all stores will be summed and we take a look at the performance of a prediction for each shoe type. In this way, each time series contains enough data for making a prediction. Hereafter a comparison between the performance of both models on entrepreneurial basis will be given. This give insight in how both models fulfill their target when less data is available. Lastly, a comparison between the computation times of both models on both levels will be given.

5.1.1 Shoe type specific forecast

Both models will be executed for 96 time series. This is equal to all possible combinations in category (men, women, children and other) and shoe type (24 in total). For example: 'Men Sneakers' and 'Women Sneakers' are two individual time series. For both models, we will take a look at the RMSE and MAE. A smaller RMSE and MAE indicates a more accurate prediction. The difference of the RMSE and MAE with respect to the naive model has been calculated in percentages for each time series individually. Moreover the MAEP is displayed in the same subdivision. The equations for the used performance measures can be found in section 4.4.

In table 4 the performance with respect to the naive model is subdivided in different percentual classes. The number of time series which belong to a specific class of performance can be found in table 4. For the time series which show an impairment with respect to the naive model, only a distinction between -10% to 0% and smaller than -10% has been made. This distinction is made because a small deterioration of less than 10% does not necessarily mean that the model is inappropriate. The same reasoning is used for the MAEP with the addition of 100% to 110% and bigger than 110%.

Table 4: Number of time series on shoe type level which show an improvement with respect to the naive model and the MAEP subdivided in different percentual improvement classes

	SVR			*LightGBM*		
	RMSE	**MAE**	**MAEP**	**RMSE**	**MAE**	**MAEP**
Smaller than -10%	57	65	-	26	26	-
-10% to 0%	10	12	-	14	20	-
0% to 10%	15	9	0	21	16	0
10% to 20%	7	4	0	13	15	0
20% to 30%	3	2	0	10	12	0
30% to 40%	1	1	0	6	3	0
40% to 50%	2	1	0	4	1	3
50% to 60%	0	0	2	1	2	1
60% to 70%	0	1	1	1	1	1
70% to 80%	0	0	1	0	0	4
80% to 90%	0	0	4	0	0	2
90% to 100%	1	1	1	0	0	0
100% to 110%	-	-	8	-	-	5
Bigger than 110%	-	-	79 (82%)	-	-	80 (83%)
Total improved	**29** (30%)	**19** (20%)	-	**56** (58%)	**50** (52%)	-

Table 4 shows that LightGBM scores better for this specific configuration than SVR regarding the RMSE and MAE decrease in comparison with the naive model.

For 30% of the time series, the RMSE of the SVR shows an improvement with respect to the naive model. As regards the MAE of the SVR in contrast to the naive model, only 20% of the time series shows an improvement. One of the time series shows an improvement of 100%. This time series can be found in the appendix. Because of only one outlier in the training data and all other data equal to zero, the SVR makes a zero prediction for all weeks, while the naive based forecast shows again the outlier in the same week. This outlier does not appear in the actual data of the same week.

LightGBM shows an improvement for more than 50% of the time series as compared with the naive model for both performance measures.

As mentioned in section 4.4, the MAE is more robust to outliers. The predicted aberrant observations for the naive model are the same as the year before, while the chance of an outlier exactly the same week as the year before is small. Therefore, naive performs better regarding the

MAE, in other words the improvement in comparison with the naive model with respect to the RMSE is higher. This difference is bigger for SVR than for LightGBM. We will come back to this in section 5.2.1 and 5.3.1 when we take a look at the behaviour of both models regarding outliers.

Concerning the MAEP, for both the SVR and lightGBM the MAE is bigger than the mean sales for 82% and 83% of the time series respectively. This is because the sales of a specific shoe type can vary a lot between weeks. For example when the MAE is high due to some missed peaks during high sales and the average sales is low because the shoes are only sold during some weeks of the year, the MAEP is very high. Regarding this performance measure, no big differences between the SVR and LightGBM can be found. The LightGBM has a few time series with a smaller MAEP than the SVR.

5.1.2 Entrepreneurial specific forecast

For the more specific prediction on entrepreneurial basis per shoe type the same measures are used to calculate the performance as in section 5.1.1. Now 1825 forecasts have been made at a time using both the SVR and LightGBM. The outcome can be found in table 5.

Table 5: Number of time series on entrepreneurial level which show an improvement with respect to the naive model and the MAEP subdivided in different percentual improvement classes

	SVR			*LightGBM*		
	RMSE	**MAE**	**MAEP**	**RMSE**	**MAE**	**MAEP**
Smaller than -10%	1265	1311	-	354	501	-
-10% to 0%	179	172	-	209	264	-
0% to 10%	153	125	24	318	298	0
10% to 20%	79	76	9	294	332	2
20% to 30%	42	43	12	275	230	5
30% to 40%	21	22	14	168	101	59
40% to 50%	22	19	75	85	55	125
50% to 60%	14	11	75	48	30	182
60% to 70%	7	6	86	31	8	176
70% to 80%	7	12	102	28	4	155
80% to 90%	11	9	110	12	0	150
90% to 100%	25	19	123	3	2	98
100% to 110%	-	-	201	-	-	183
Bigger than 110%	-	-	994 (54%)	-	-	690 (38%)
Total improved	**381** (21%)	**342** (19%)	-	**1262** (69%)	**1060** (58%)	-

If we take a look at the total number of time series which show an improvement in comparison with the naive model for this configuration, the lightGBM scores again best for both measures. This also holds regarding the MAEP.

When it comes to the SVR, just 20% of the times series show a decrease in RMSE and MAE relative to the naive model. The difference between those two performance measures is somewhat decreased. The impact of aberrant observations is smaller for this configuration of forecasting in comparison with the forecast per shoe type. Because it is a forecast over all retailers subjected to different shoe types, some time series only contain one or a few observations, which results in a bad performance of the naive model. Therefore, more time series can be found in the higher improvement classes.

If we now take a look at the performance of LightGBM with respect to the naive model, it shows an improvement for almost 60% of the time series regarding the MAE and for 70% of the time series with respect tot the RMSE. A relatively big difference between the RMSE and MAE can be found. Because of fewer data points per time series the data is noisier. A smaller amount of time series can be found in the higher improvement classes concerning the MAE.

When we take a look at the MAEP, we see a great improvement for both models in comparison with the MAEP of the shoe type specific forecasts. The sales of each shoe type summed over all retailers, are much higher, than for each retailers individually. Due to the more extreme values of sales which occur for each shoe type, the MAEP is higher. Based on this performance measure, LightGBM again outperforms the SVR. Where the MAE of the LGBM predictions for 38% of the time series is bigger than the mean sales, is this the case for 54% of the time series for SVR. However, SVR shows more time series in the lower percentual classes. The LGBM searches for a good forecast in general, where the SVR can produce highly accurate predictions for some time series.

When both configurations (shoe types in general and focused on each retailer separately) are compared using the same model, no extreme differences can be found as regards the improvement with respect to the naive model. Where the SVR performs slightly better for the prediction for each shoe type compared to the prediction on shoe type for each entrepreneur, holds the opposite for LightGBM: it performs slightly better on entrepreneurial basis compared to a forecast for each shoe type with a summation over all retailers. The MAEP is for both models decreased when looking at the entrepreneurial based forecast because of the smaller absolute fluctuations and outliers in sales.

5.1.3 Computation time

Every week new data becomes available and the model has to be updated and produce a new prediction for all coming 52 weeks. Therefore running the prediction algorithm should not last forever. All predictions for this research are performed using a HP Windows 10 Pro laptop with an Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz processor. The data generation for the ShoeMonitor will be done automatically on an external server with enough CPUs. In this way, not every retailer has to run the model independently.

The computation times of both models are compared in table 6

Table 6: Computation time of both models on both levels

	SVR	**LightGBM**
Entrepreneurial Forecast	31:03 h	07:21 h
Shoe type Forecast	06:38 h	00:46 h

Regarding computational feasibility, LightGBM strongly rises above SVR. Although LightGBM contains more options in its grid search, the implementation of both GOSS and EFB which are combined in LightGBM provide a quick way to execute this grid search. We can conclude that leaf-wise growing tree-structure has a great positive impact on the computation time.

5.2 SVR

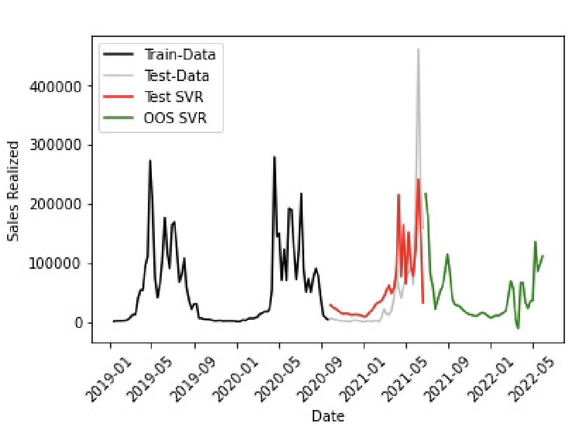
Now the performance of both models has been compared, we are going to take a look at each model separately. This will give understanding in what type of prediction SVR is more appropriate and what are its pitfalls. A few time series will be highlighted to get acquainted with common patterns. For the following sections the same motives have been used to select the time series. First, the time series which shows the best performance regarding this task for its specific configuration will be shown (for example: best prediction with SVR on shoe type specific level). The next chosen time series shows the time series which has the best performance using the other model. This provides insight into whether it is a generally well-predictable time series or whether the other model really performs better because of certain aspects (for example trend). Hereafter the same three plots will be displayed for both models. One time series which shows strong seasonality, a time series with a big outlier and the last one which shows a clear trend. All 6 performance measures are displayed for completeness. A standalone MAE or RMSE does not say much if you don't know the range of the values. Therefore the mean sales of the test set ($\bar{\boldsymbol{y}}_{\boldsymbol{test}}$) as well as the improvement of the RMSE and MAE with respect to the naive model are displayed as well. Furthermore, all different time series are numbered for anonymity. This is the index number of that time series. Because of the big difference in size of the data, both the forecasts on entrepreneurial basis and shoe type basis will be evaluated.

5.2.1 Shoe type specific forecast

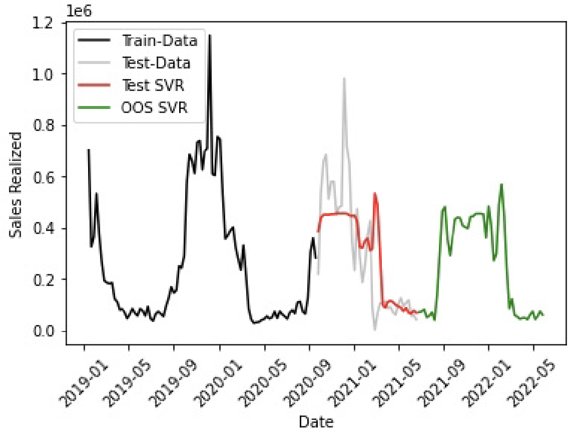
In table 7 an overview of the error measures for this type of forecast is given for 5 time series. The equations used for the mentioned performance measures can be found in section 4.4 (equations 14, 17, 15, 18, 20 and 21). In figure 12 all mentioned time series are plotted for the whole period including train-, test-data and the out-of-sample prediction. The out of sample forecast is also plotted to see whether the forecast one year ahead does not show strange patterns based on intuition created with knowledge on the historical data. This gives also some knowledge of overfitting. If the out-of-sample prediction is the same as the prediction of the test-data, the model is probably overfitted.

Table 7: Overview of mean sales and performance measures for the highlighted time series prediction on shoe type level using SVR

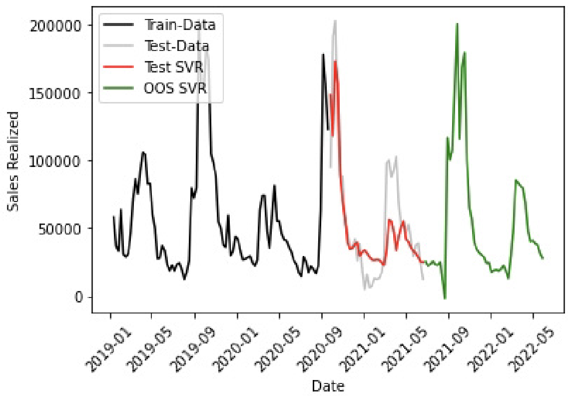
Index	$\bar{y}_{test}$	**RMSE**	**% change**	**MAE**	**% change**	**R^2**	**R^2 adjusted**
36	14924,10	57627,28	12%	36037,68	34%	0.55	0.47
47	271403,93	176483,18	14%	116540,72	4%	0.47	0.40
7	80832,60	27096,67	-23%	19035,72	-22%	0.65	0.42
32	167,30	54237,35	1%	24911,61	-8%	-0.01	-0.31
24	490,26	1219,62	10%	1064,58	-2%	-1.44	-2.87



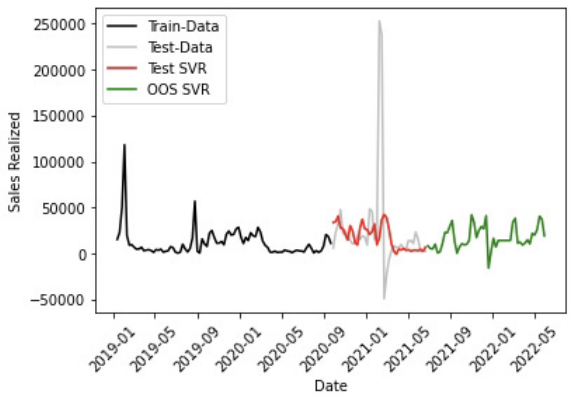
(a) Plot of time series 36



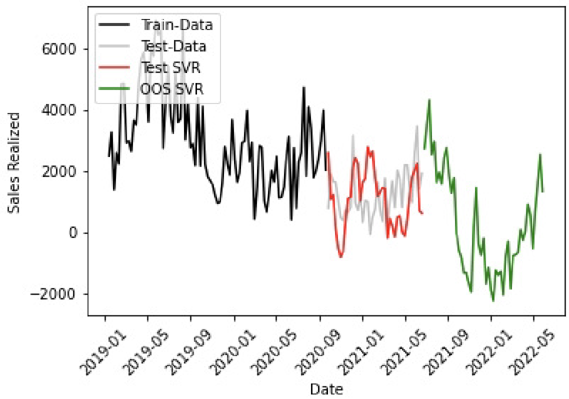
(b) Plot of time series 47



(c) Plot of time series 7



(d) Plot of time series 32



(e) Plot of time series 24

Figure 11: Highlighted time series of SVR forecast for different shoe types

First, we will have a look at figure 11a, the plot of time series number 36. This time series shows the overall best execution for this task. This prediction shows both a good fit regarding the test data and an improvement with respect to the naive model (see table 7). The times series shows clear seasonality which is captured in the model prediction as well.

The next chosen time series is number 47. This is the time series with the best performance using LightGBM. The time series shows both seasonality and a downward trend (figure 11b). Around December a peak can be found which is not predicted with SVR. Because of the small amount of data this peak only occurs once and is not (yet) included in the model. Furthermore, the model does not forecast a decrease of sales in its out-of-sample forecast. Thus, the SVR does not recognize the trend. This also has to do with the fact that little historical data is available yet. The fit of LightGBM is better, thus LightGBM can produce a more accurate prediction using less data than SVR.

The prediction of times series number 7 (figure 11c) shows a better fit when it comes to the R^2 and R^2_{adj} compared to time series 36 (table 7). However, when compared to the naive model, this prediction has both a higher RMSE as well as MAE. Because of the relatively small amount of years of historical data, it detected a downward trend for the peak sales around April in 2019 and 2020, while this trend did not continue in 2021. Therefore, the naive model prediction is more accurate for this specific time series. Even though the prediction around April 2021 is too low, the model fit is adjusted for the forecast of April 2022 and the downward trend is not visible anymore.

Subsequently, will have a look at time series 32 (figure 11d). This time series shows the sales of rain and snow boots. This plot shows two extreme outliers, one upwards and one downwards. We can assume that it snowed that week or snow was forecasted for the coming week. This is a random aberrant observation and therefore the model was not able to predict it. It shows us that when a outlier only happens once in a specific week, it will not necessarily appear in the prediction of the coming year.

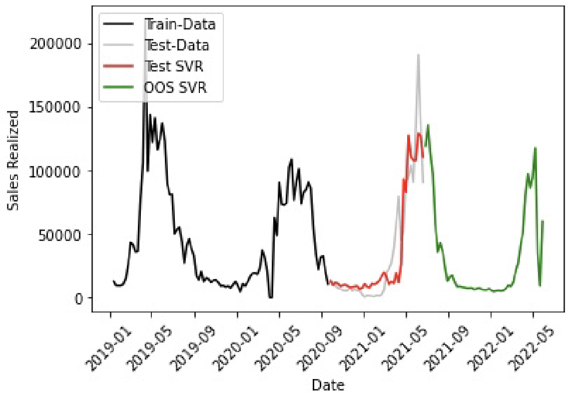
The last highlighted time series for this set-up is shown in figure 11e. This time series shows a clear downward trend, which is visible in all years. The model is able to make a prediction which takes this trend into account. Moreover, the data for this time series is very noisy and thus the provided forecast does not have high $R^2_{(adj)}$. Concerning the percentual changes with the naive model, it shows an improvement regarding the RMSE and a decrease regarding the MAE. This is due to the noisiness of provided the data and the robustness of the MAE. The downward trend in shoe sales will not always continue. Shoes have a certain lifespan and therefore shoes will always be sold or a shift in shoe types has taken place where a certain type disappears and its sales becomes zero. So it is unlikely that sales will continue to fall for time series 24. The model will adjust the trend with the weekly update by the addition of actual data.

5.2.2 Entrepreneurial specific forecast

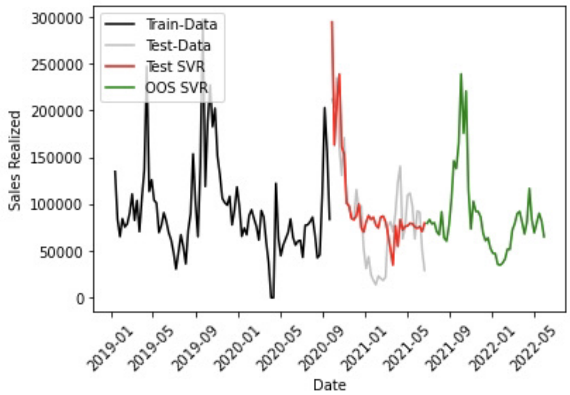
In this subsection we will take a look at the performance of the prediction for each retailer individually. Therefore more predictions are made with a smaller size and less data points for some time series. Again 5 time series will be highlighted to give an overall interpretation of SVR performance on this level. The same reasoning as mentioned in the beginning of section 5.2 is used to select the time series shown.

Table 8: Overview of mean sales and performance measures for the highlighted time series prediction on retailer level using SVR

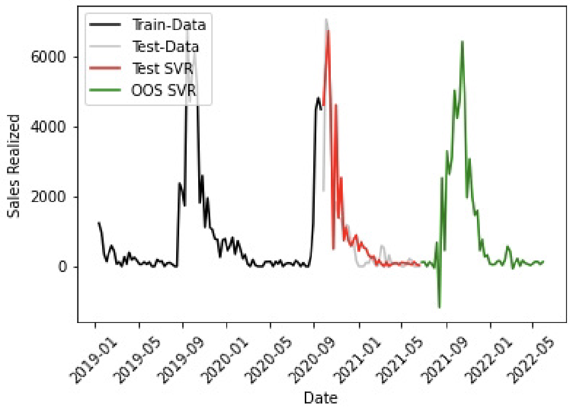
Index	$\bar{y}_{test}$	**RMSE**	**% change**	**MAE**	**% change**	**R^2**	**R^2 adjusted**
476	35608,83	20760,90	19%	13731,58	12%	0.80	0.74
4	87772,03	41205,51	20%	33488,30	15%	0.43	0.33
242	860,98	720,36	13%	471,58	17%	0.80	0.72
783	3536,19	4140,07	52%	2516,03	-26%	-4.16	-6.78
594	479,73	1541,22	-65%	1274,86	-76%	-15.16	-25.70



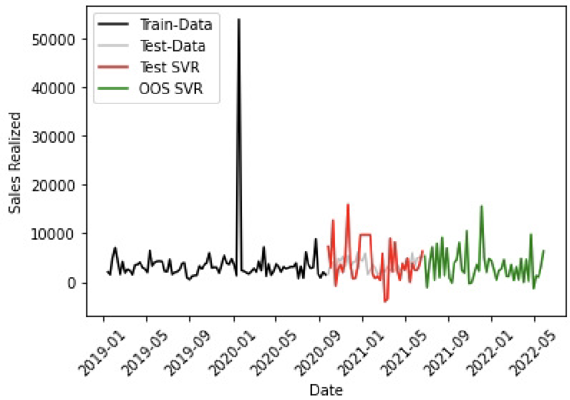
(a) Plot of time series 476



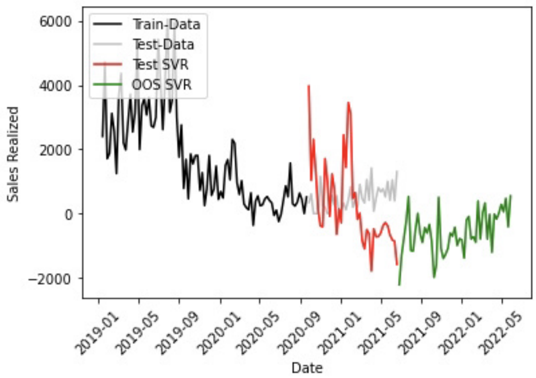
(b) Plot of time series 4



(c) Plot of time series 242



(d) Plot of time series 783



(e) Plot of time series 594

Figure 12: Highlighted time series of SVR forecast on retailer specific basis

If we take a look at the time series number 476 (table 8 and figure 12a) we see a big improvement as regards the R^2=0.88 when compared to the best time series forecasted on shoe type level (R^2 = 0.55). The forecast of this time series shows an improvement with respect to the naive based forecast as well. This forecast shows the most accurate fit of all time series for SVR. The series show clear seasonality which also arises in the forecast. Although for these time series filtering on retailers has taken place, the overall sales is relatively high. This is thus the the time series of a big shoe retailer. This means the data still has a significant size.

Time series 4 (figure 12b) is the time series with the best performance using LightGBM. However, the prediction using SVR does also show an improvement and a fine fit (8), but not as good as LightGBM. The time series of this retailer focused on a specific shoe type has an average sales per week of approximately €100000 and some upward and downward peaks every year. The SVR is not able to generate both peaks in the test-data. The pattern of first a peak sales than a fall in sales is visible in the out-of-sample data though. Therefore, we can conclude the SVR will include a returning peak and trough sales if it occurs two years in a row.

The division in a general shoe type forecast and shoe type per retailer is made because of the difference in size of data. Time series 242 (figure 12c) shows a series with a smaller size. For this time series an accurate forecast can be made as well, based on all performance measures, thus the SVR can make a precise prediction, even if the size of the data is smaller. What is further remarkable, is the negative predicted sales around September 2021. The model predicts more shoes will be returned to the store. The model recognizes a pattern the weeks before and therefore predicts a fall in sales which even results in a negative sales.

The next time series shows one extreme outlier (figure 12d). From this we conclude that SVR does not incorporate the outlier, if it only occurs once. Because of this outlier, the RMSE of the prediction shows a great improvement with respect to the naive RMSE, but this improvement does not hold for the MAE. Because of the outlier, noisier data is predicted which cannot be found (or in a lesser extent) in the actual data.

The last highlighted plot is the plot of time series 594 (figure 12e), which is highlighted for its trend. The SVR detects the trend and makes the prediction based on this. Therefore, the sales end below zero. This is often not the case for a long period because of the ending lifespan of a shoe. However, the model adjusts its own fit in the out-of-sample forecast. A small upward trend can be found.

To summarize, the SVR is able to produce good predictions for some time series on both levels. The pitfalls are mainly when a trend is present. Although it recognizes the trend, it does not ensures the right fit. Shoes are products with a lifespan which we will always need. Therefore the downward trend will not last forever. As regards seasonality, the SVR is able to recognize is and apply it to the prediction.

5.3 LightGBM

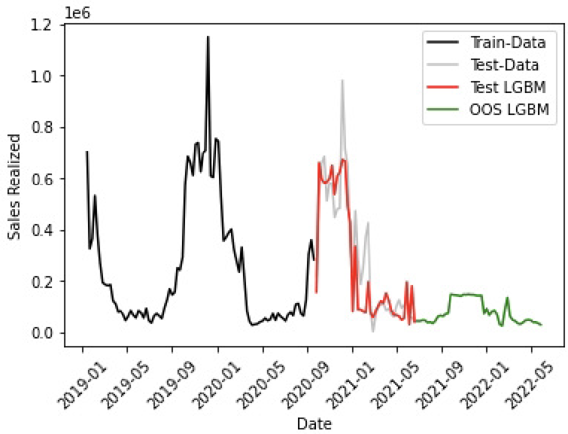
Let us now take a closer look at the results and common patterns of LightGBM. For this subsection a distinction has been made again between the shoe type specific forecast and the fit for each shoe type focused on the retailers. Furthermore, the same performance measures are used as for SVR and the same time series are highlighted as in sections 5.2.1 and 5.2.2.

5.3.1 Shoe type specific forecast

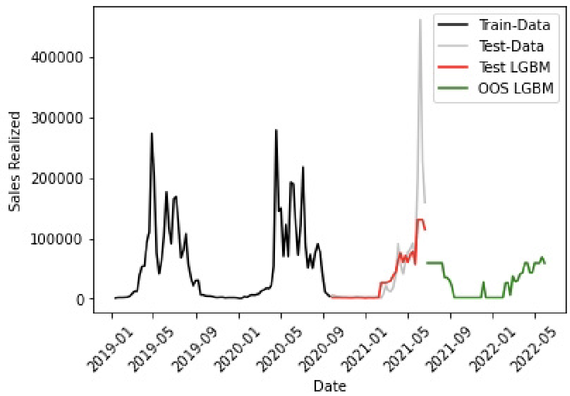
In table 9, a summary is given for the 5 highlighted time series with its performance measures and the mean sales of the test data ($\bar{y}_{test}$). The corresponding plots are given in figure 13.

Table 9: Overview of the mean sales and performance measures for the highlighted time series prediction on shoe type level using LightGBM

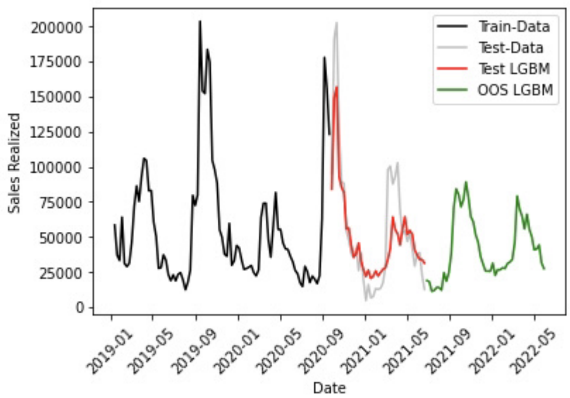
Index	$\bar{y}_{test}$	**RMSE**	**% change**	**MAE**	**% change**	**R^2**	**R^2 adjusted**
47	271403,93	115052,77	25%	87003,07	22%	0.78	0.73
36	14924,10	57441,79	12%	20041,67	25%	0.56	0.45
7	80832,60	22996,00	-5%	16227,03	-5%	0.75	0.64
32	167,30	54868,62	-1%	22897,59	-1%	0.0	-0.4
24	490,26	813,57	40%	671,36	35%	-0.08	-0.80



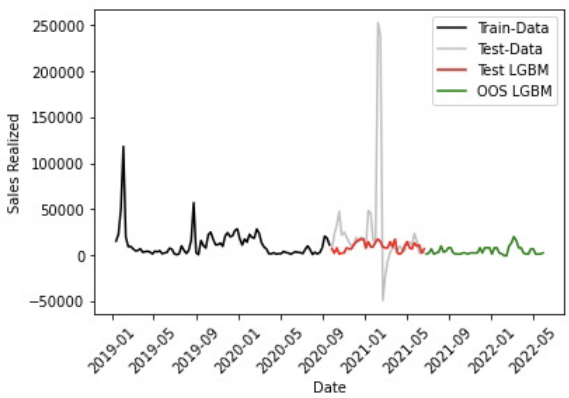
(a) Plot of timeseries 47



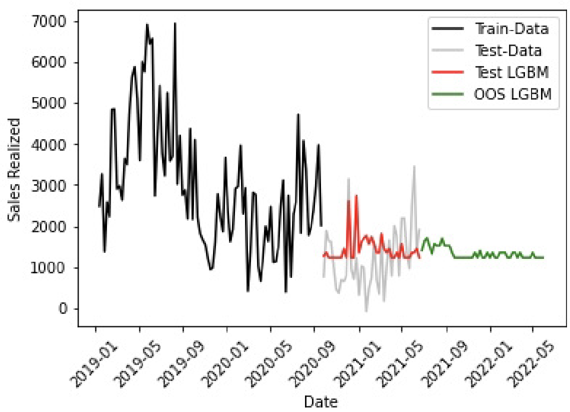
(b) Plot of time series 36



(c) Plot of timeseries 7



(d) Plot of timeseries 32



(e) Plot of timeseries 24

Figure 13: Highlighted time series of LGBM forecast on shoe type specific basis

The first mentioned time series, number 47, shows both the largest improvement with respect to the naive model, as well as the best fit regarding the $R^2_{(adj)}$. The time series show clear seasonality as well as a downward trend. The fit regarding the test data is better than for SVR, because of the higher prediction of sales around November 2020. However, it does not forecast the high sales around November 2021. Therefore this fit could be improved with more years of historical data.

The time series with the best fit regarding the SVR model, time series 36 (figure 13b), shows almost the same accuracy in relation to the SVR model (table 9 and table 7). The out-of-sample prediction for the SVR is much steeper in the last predicted weeks than for the LightGBM. Looking at the historical patters, the SVR model seems more logical. Although time series 7 shows a good fit regarding the $R^2_{(adj)}$ (even better than the fit of time series 7 with SVR), the out-of-sample prediction of the SVR looks more logical. Because of the small amount of data the bigger outlier around September each year is not fully taken into the model prediction of LightGBM. Because of the occurrence of two peaks each year, the LightGBM predicts a kind of average peak twice every year instead of a bigger and smaller aberrant observation. In case of one big outlier as in figure 13d, the LightGBM does not add this directly to the model and all other predictions are less affected by this outlier than for the prediction using SVR (figure 11d).

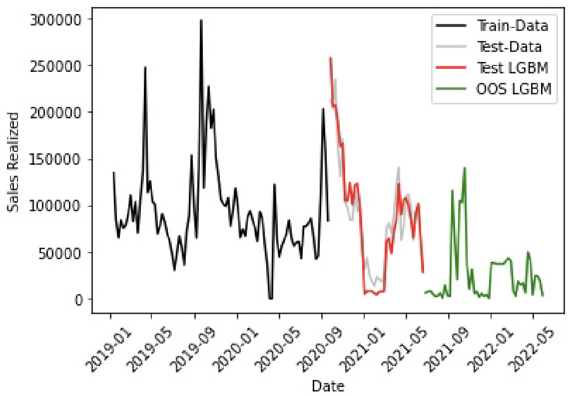
Let us now take a look at the time series which includes a clear trend, time series 24 (figure 13e). The time series shows an overall downward trend. Because of the wave-like movement in the test-data, the LightGBM has difficulty in finding the right pattern. The advantage of this is that it does not predict a negative sales for a longer time, like the SVR did (figure 11e). However, all other patterns are diminished as well by the LightGBM fit.

5.3.2 Entrepreneurial specific forecast

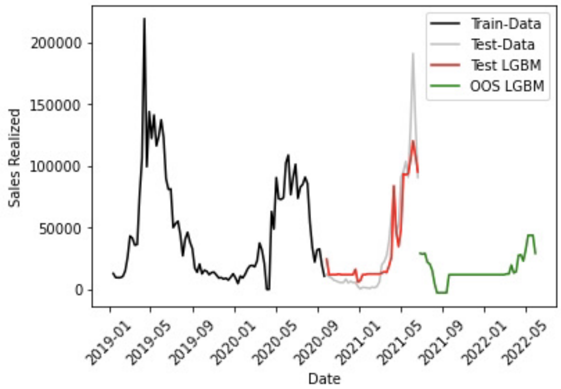
We will now have a look at the performance of the LightGBM on retailer specific level. Therefore the same 5 time series are highlighted in table 10 and figure 14.

Table 10: Overview of the mean sales and performance measures for the highlighted time series prediction on retailer level using LightGBM

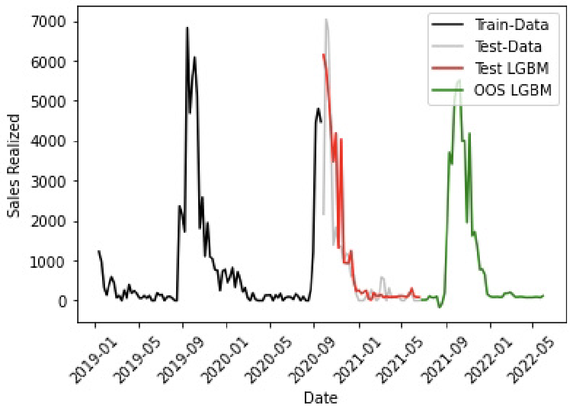
Index	$\bar{y}_{test}$	**RMSE**	**% change**	**MAE**	**% change**	**R^2**	**R^2 adjusted**
4	87772,03	20364,53	60%	16512,64	58%	0.87	0.77
476	35608,83	17568,09	31%	11851,85	23%	0.86	0.78
242	860,98	994,62	-21%	488,342	-4%	0.63	0.55
783	3536,19	2457,04	71%	1793,18	32%	-2.1	-3.3
594	479,73	648,94	30%	540,37	25%	-0.51	-1.23



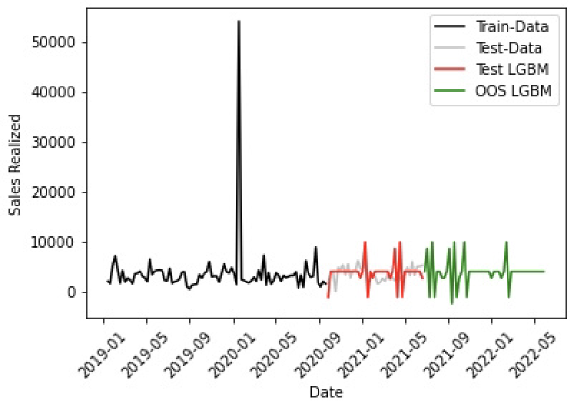
(a) Plot of time series 4



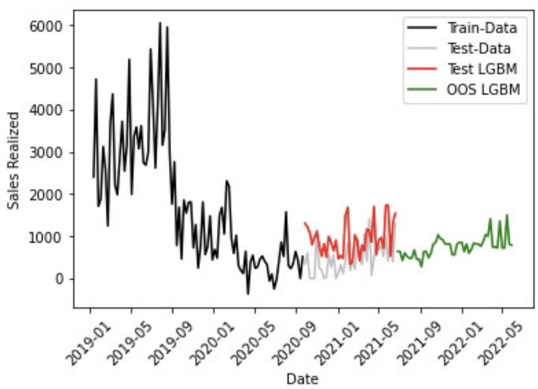
(b) Plot of time series 476



(c) Plot of time series 242



(d) plot of time series 783



(e) Plot of time series 594

Figure 14: Highlighted time series of LGBM forecast on retailer specific basis

In table 10 it can be found that the prediction of time series 4 has a high accuracy. It shows a high $R^2_{(adj)}$ and because of the downward trend (see figure 14a) a great improvement in comparison with the naive model.

Figure 14b shows the plot of the time series with the highest accuracy for SVR. For LightGBM, the forecast has a high accuracy as well, even better than for the SVR. The out-of-sample prediction of the SVR shows a increase in sales in the last weeks of the prediction, where this is not the case yet for LightGBM. More historical data is needed for LightGBM to predict the start of increase in sales more precisely. When less sales are incorporated in the data, as is the case for time series 242 (figure 14c, LGBM can make an accurate prediction as regards the R^2. The fit is a little worse when compared to the naive model (table 10), but incorporates the seasonality well. The negative spike which can be found in the forecast of this time series by SVR (figure 12c), can not be found in this forecast. If the data does not contain any big negative values, LightGBM will not forecast big negative value as well. This does not explicitly hold for the SVR model.

We will also take a look at time series 783, where the data contains one big outlier (figure 14d). This time series shows great improvement regarding the RMSE in comparison with the naive model. Because the MAE is a more robust performance measure, the improvement with respect to the naive model of the MAE is much smaller. In comparison with the SVR, the LightGBM has a higher accuracy. The predicted data is due to the outlier and no other clear patterns noisier in comparison with the actual data for both the SVR and LightGBM, where for LightGBM the effect of the aberrant observation is smaller.

The last plot will give us understanding in the performance of prediction when a trend is visible. As mentioned above, the LGBM will not predict big negative values when those do not appear in the data. This becomes clear from the plot of time series 594 as well (14e). Where the SVR predicted an ongoing negative trend, the LightGBM make a more stabilized prediction. Therefore, the fit is better than both the SVR model as well as the naive model.

In conclusion, LightGBM shows good performance with respect to seasonality and trends. The big advantage of LightGBM in contrast to SVR is that when the model is not familiar with a certain amount of sales (for example negative sales), it will not predict it. Improvement could be made when more historical data is available. Some general shoe types can contain a peak sales twice a year. Because of the lack of historic data, the model has difficulty in predicting both peaks. For those type of time series prediction using the naive model is slightly better. Regarding time series which contain an outlier, for example for the snow boots prediction, its general fit is also not very good. For these shoe types the prediction will never get very accurate because of its extreme randomness. Predicting these specific peaks is not the aim of this study. When an outlier occurs, the general fit of the model is also affected and lose track in all weeks.

Therefore, the model fit is more accurate using a naive model.

5.4 Forecast with COVID-19 adjustment

In this last subsection of the results we will have a look at the influence of COVID-19 on the sales forecast. The forecast is produced using data from January 2018 until June 2021. From March 2020, the pandemic had a great influence on the shoe sales because of customer reluctance and measures prepared by the government (see section 2.3). The already present effects on sales can be found in section 3.4. Therefore we will have a look at the effect on the prediction when no adjustment has been made, as well as the effect on the forecast if the weeks in which COVID-19 had a great impact on the sales (for example when the shops were closed) are removed from the model and the other weeks are interpolated. The weeks in which an adjustment has been made can be found in section 3.4. The adjusted week 12 and 13 of 2020 are part of the train-data, week 50 2020 until week nine of 2021 are part of the test-data. The effect is studied to get knowledge of the short- and long-term impact when a couple of weeks are affected for both SVR and LightGBM. This does not include the fact that the closing of the stores may result in peak sales when the shops are reopened. We only looked at the effect on the forecast when the reduced sales are removed and therefore will not be included as lagged variable in the forecast for next year.
The same test- and train-data split has been used. This means that 70% (89 weeks) is used for training and 30% (39 weeks) is used for testing.

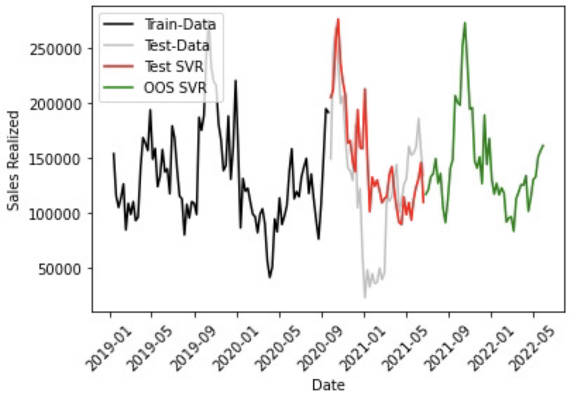
5.4.1 SVR

In table 11 the number of time series which show an improvement with respect to the naive model are displayed. A distinction has been made on both the performance measures, as well as if a COVID-19 adjustment has been done or not.

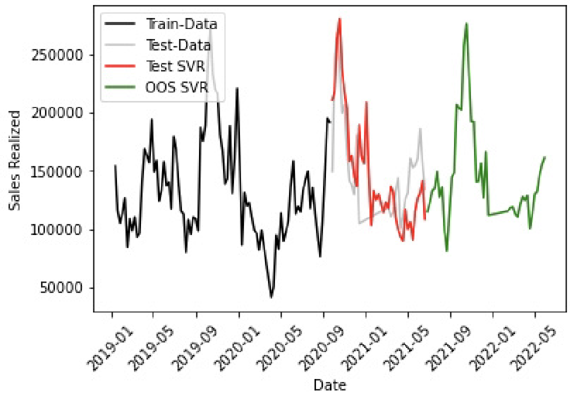
Table 11: Number of time series which show a decrease in RMSE and MAE for SVR with respect to the naive model. A subdivision had been made in time series with and without COVID-19 adjustments.

	No adjustment	**COVID-19 adjustment**
RMSE	29	33
MAE	19	23

A small increase can be found in the number of time series which show an improvement for both the RMSE and MAE.



(a) Without COVID-19 adjustment



(b) With COVID-19 adjustment

Figure 15: Plot of a time series which show a great improvement after the COVID-19 adjustment

In figure 15 a time series has been highlighted which show clear effect of the COVID-19 virus. In the test-data a clear trough in sales can be found around the end of 2020 and start of 2021. As a result of these trough sales, a small peak can be found in the following months. As mentioned before, only the weeks with a trough in sales are removed from the model. The following peak sales are not adjusted.

Figure 15a shows the data and prediction of the time series when no adjustment has been made. Figure 15b shows the plot when the weeks which are affected by the virus are removed from the model. No big differences between those two forecasts can be found. The impact of COVID-19 on the prediction with SVR is not striking. This is in line with the fact that one aberrant observation is not enough to include it in the model as found in section 5.2.

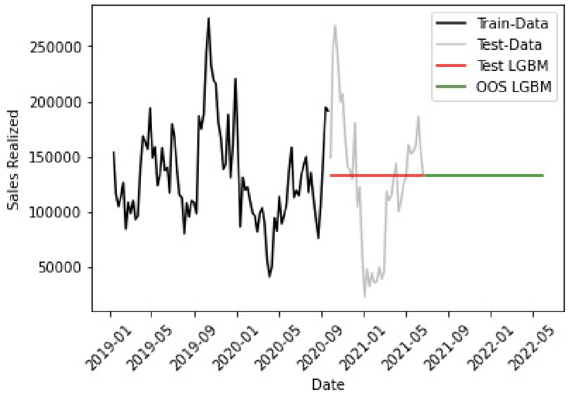
5.4.2 LightGBM

Now the same comparisons will be made as for the SVR model. In table 12 a short summary can be found.

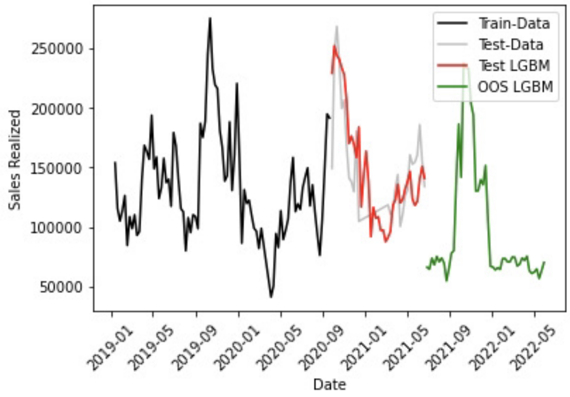
Table 12: Number of time series which show a decrease in RMSE and MAE for LGBM with respect to the naive model. A subdivision had been made in time series with and without COVID-19 adjustments.

	No adjustment	**COVID-19 adjustment**
RMSE	56	55
MAE	50	55

Again, no big difference can be found in the number of time series which show an improvement between the COVID-19 adjusted series and the series without adjustment. When the MAE is used as performance measure we see a small increase when the COVID-19 adjustment has taken place.



(a) Without COVID-19 adjustment



(b) With COVID-19 adjustment

Figure 16: Plot of a time series which show a great improvement after the COVID-19 adjustment

Figure 16 shows the same time series as used as an example in section 5.4.1. For LightGBM the adjustment ensures an improvement in terms of fit. Where initially the model had difficulty finding the right fit and thus predicted a straight line, is this not the case after the COVID-19 adjustment (figure 16b). The big fall of sales around January 2021 is not visible anymore after the adjustment which makes it possible for the LightGBM to find a pattern and thereby a fit with an R^2 of 0.46.

In conclusion, both the SVR and LightGBM do not show great increase in number of time series which show an improvement with respect to the naive model after some adjustments of the data due to the Coronavirus. However, the forecast with LightGBM is more affected by this decrease in sales. If the data contains just one outlier, it will not affect the fit of the time series by LightGBM (see section 5.3), but if this aberrant observation lasts for a couple of weeks (as is the case during the pandemic), the LightGBM fit gets affected. The adjustments make it possible for LGBM to find a pattern and base a prediction on this in combination with all other variables.

6 Discussion & Conclusion

In this research, we have examined the predictability of shoe sales for multiple shoe stores in the Netherlands. Both big retailers, as smaller retailers are covered by the forecast. Sales forecasting is a popular research topic. Having an accurate prediction can help both manufacturer and retailer, but the occurrence of multiple unforeseen aspects which influence the sales (for example fashion trends) makes it challenging. To get understanding about the usefulness of the forecast, the predictions are compared to the naive model because most retailers nowadays base their procurement on their sales of last year. To make the prediction two models were introduced: the support vector regression model and the relatively new and popular LightGBM. Both models are used for making a forecast for the coming 52 weeks. Therefore, a distinction has been made between a more high over prediction on shoe type basis and a forecast focused on the sales of each retailer individually. This forecast is also subdivided into different shoe types. Because of the abnormal situation COVID-19 which has a great impact on sales as well, some primary research has been executed to get feeling of its impact on forecasting. The weeks in which the sales are highly affected by the virus, because of for example forced closing of stores, are removed from the model. This provides knowledge of the effect of trough sales on forecasting.

Where the SVR model prediction on shoe type basis only outperformed the naive model for approximately 20% a 30% of the time series (based on MAE and RMSE respectively), was this the case for LightGBM for more than 50%. On retailers specific basis, the LGBM scores even better with an improvement for almost 60% (even almost 70% w.r.t. the MAE) of the time series, where SVR still improved for 20% of the time series.
Concerning the MAEP, an improvement can be found when forecasting on retailer specific level because of the smaller differences in sales over the year. As regards this performance measure, the LGBM outperformed the SVR as well. Due to great differences in sales of certain shoe types during a year, the MAEP shows high values for many time series.
Besides the difference in accuracy, a weekly computation time of almost 38 hours using SVR to produce a prediction on both levels, is far from desirable. In contrast, the computation time of producing both prediction levels using LGBM is 'only' 8 hours which is feasible for a forecast which is updated every week.

Literature showed promising results regarding SVR for retail forecasting, although for this research the naive model showed better performance for most time series. This is partly due to the lack of lots of historic data. Earlier accurate produced predictions using SVR in sales forecasting used more historic data ((Wen et al., 2014) and (Yangl et al., 2007)). SVR does not incorporates specific patterns or aberrant observations if they only occur once. Thus more historic data ensures a better fit regarding for example seasonality. Moreover, a subdivision in multiple characteristics of shoe sales, such as seasonality and outliers, did not result in a general pitfall or general pitfalls of the SVR regarding its performance. However regarding trend, improvement could be made. Due to the relatively small amount of data in number of weeks, the SVR quickly recognizes a trend which is only visible in the train-data. The continuing of this trend causes a lower accuracy regarding the test-data. Another shortcoming of the trend recognition of SVR is that it can produce a forecast of only negative sales, while this is not logical given the lifespan of shoes.

Concerning LightGBM, this research confirms the findings of the M5 forecasting competition: the algorithm can produce an accurate forecast in a relatively short period of time (Makridakis et al., 2020a). The leaf-wise tree growth feature of this gradient boosting algorithm has a positive impact on the computation time. The same subdivision has been made as regards the characteristics (trend, seasonality and outliers). For LGBM, no clear driver of general misfit has been found. The overall fit of LGBM is better as well as the reaction on a trend in comparison with SVR. Where the SVR continues downward trends to negative sales, is this not the case for LGBM. LGBM will only predict values based on the historic sales interval. It will therefore not forecast negative values of sales which have never been present before. This has to do with the fact that it is a tree based algorithm, where the historic sales are used to create the branches and thus the outcome is between known margins of the produced branches for that time series. As an exception, some time series contain peak sales more than one moment a year. In this case the LGBM shows difficulty in forecasting all peaks. The lack of more years of historic data causes a misfit in this case of the prediction.

To make these predictions, the same variables are used for both models on both levels. The added external variables are: school holidays with its spreading, normal holidays and seasons. For the highlighted time series the size of difference between the R^2 and $R^2_{(adj)}$ can differ a lot. Some time series show a big difference between both performance measures, others only a few tens. This means that not all variables have the same impact for every shoe type. Because of the importance of computational feasibility, the choice has been made to not include a grid search on which variables to add for each time series. The addition of these external variables show promising results for multiple time series and are therefore included for all time series. Further research on the impact of all variables per time series is useful.

During multiple weeks of the given historical data, we lived in the COVID-19 pandemic. This pandemic has a great impact on the sales due to for example the forced closing of stores. The effect of the pandemic on sales forecasting will be long lasting because the predictions are mostly based on historic data and patterns. Research has been executed on the effect of supply chain disruptions and the spread of the virus. However, there is a lack in research of the effect on sales forecasting. A small study has been done to measure the effect of the virus on both models for this specific prediction task.

Regarding the number of time series which show an improvement with respect to the naive model, no big differences can be found for both models between the adjusted time series and the time series without adjustment. However when we take a look at a time series which shows clear effect of the pandemic, a distinction between both models can be found. The SVR is not highly influenced by the aberrant observations caused by the pandemic, while the LGBM cannot find the right pattern because of the polluted data. It therefore forecasts a straight line. The COVID-19 adjustments (removing weeks affected by the virus) causes the model to find a pattern and uses this for producing the out-of-sample forecast.

In this research, only the trough sales are removed from the model. However, as a result of the trough sales, a peak sales occurs. Shoes are essential products and therefore this reaction is present in the data. Further research could be done to improve the forecast based on both the trough and peak sales. Therefore, information about the mean sales of shoes should be included in the model.

As mentioned above for both models, improvement in accuracy will be accomplished when more historic data is available. This forecast is solely based on data from January 2018 till July 2021, where the data of 2018 is only used as a lagged variable. Therefore both models have difficulty in recognizing patterns which return every year.

Besides the amount of data, improvement could be made on different aspects of algorithms and data as well. For example, now all predictions are made at once. Because of the big amount of time series to make a prediction for, none of it is adapted individually. The grid search takes a wide variability in its coefficients, but some of the time series are divergent and need a specific setting.

More research can be done on the difference of making every week a new prediction for the whole data set or making a prediction based only on the addition of one week of new data. What could be thought off is running the full data set only once a month and for the other weeks using the same fit as the week before but than also predicting one week extra. This might have a positive impact on the computation time of both models.

Furthermore, some characteristics on the stores could be included as well to produce a more accurate forecast. For example information about the location (city centre or in the suburb) could be included or a multivariate prediction based on information about the sales of stores in the neighbourhood as well could be produced.

Another point of attention for weekly based forecasting is the occurrence of leap years. For those forecasts, all years are assumed to have 52 weeks, while leap years contain 53 weeks. Further research can be done in measuring the effect of leap years and how the capture these effects in the model.

Although in the data brands are available, in these approaches the brands are unused. Future research could be done in the field of brands and clustering to add these to the model as an extra variable to provide a more precise forecast. Besides some unused but available aspects in the data should have a review for its impact on the forecast, other external variables could be taken into account as well. What could be thought of is data on the life cycle of shoes.

References

Huber, J. and Stuckenschmidt, H. (2020). Daily retail demand forecasting using machine learning with emphasis on calendric special days. *International Journal of Forecasting*, 36(4):1420–1438.

Ivanov, D. (2020). Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case. *Transportation Research Part E: Logistics and Transportation Review*, 136:101922.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Khakpour, A. (2020). Using machine learning and big data in sales forecasting for production and retail.

Krishna, A., V, A., Aich, A., and Hegde, C. (2018). Sales-forecasting of Retail Stores using Machine Learning Techniques. In *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, pages 160–166, Bengaluru, India. IEEE.

Kumar, A., Shankar, R., and Aljohani, N. R. (2020). A big data driven framework for demand-driven forecasting with effects of marketing-mix variables. *Industrial Marketing Management*, 90:493–507.

Lau, H.-S. (1980). The newsboy problem under alternative optimization objectives. *Journal of the Operational Research Society*, 31(6):525–535.

Makridakis, Spiliotis, S. ., Assimakopoulos, E. ., and Vassilis (2020a). The M5 Accuracy competition: Results, findings and conclusions.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLOS ONE*, 13(3):e0194889.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2020b). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1):54–74.

Nikolopoulos, K., Punia, S., Schäfers, A., Tsinopoulos, C., and Vasilakis, C. (2021). Forecasting and planning during a pandemic: COVID-19 growth rates, supply chain disruptions, and governmental decisions. *European Journal of Operational Research*, 290(1):99–115.

Pillo, G. D., Latorre, V., Lucidi, S., and Procacci, E. (2016). An application of support vector machines to sales forecasting under promotions. *4OR*, 14(3):309–325.

Ramos, P., Santos, N., and Rebelo, R. (2015). Performance of state space and arima models for consumer retail sales forecasting. *Robotics and Computer-Integrated Manufacturing*, 34:151–163.

Singh, A. P., Gaur, M. K., KumarKasdekar, D., and Agrawal, S. (2015). A study of time series model for forecasting of boot in shoe industry. *International Journal of Hybrid Information Technology*, 8(8):143–152.

Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222.

Sun, Z.-L., Choi, T.-M., Au, K.-F., and Yu, Y. (2008). Sales forecasting using extreme learning machine with applications in fashion retailing. *Decision Support Systems*, 46(1):411–419.

Thomassey, S. (2013). Sales forecasting in apparel and fashion industry: A review. In *Intelligent Fashion Forecasting Systems: Models and Applications*, pages 9–27. Springer Berlin Heidelberg.

Wen, Q., Mu, W., Sun, L., Hua, S., and Zhou, Z. (2014). Daily sales forecasting for grapes by support vector machine. In Li, D. and Chen, Y., editors, *Computer and Computing Technologies in Agriculture VII*, pages 351–360, Berlin, Heidelberg. Springer Berlin Heidelberg.

Wu, Q. (2009). The forecasting model based on wavelet support vector machine. *Expert Systems with Applications*, 36(4):7604–7610.

Wu, T., Blackhurst, J., and O'grady, P. (2007). Methodology for supply chain disruption analysis. *International Journal of Production Research*, 45(7):1665–1682.

Yangl, Y., Fulil, R., Huiyou, C., and Zhijiaol, X. (2007). SVR mathematical model and methods for sale prediction. *Journal of Systems Engineering and Electronics*, 18(4):769–773.

Yu, X., Qi, Z., and Zhao, Y. (2013). Support vector regression for newspaper/magazine sales forecasting. *Procedia Computer Science*, 17:1055–1062.

A Appendix

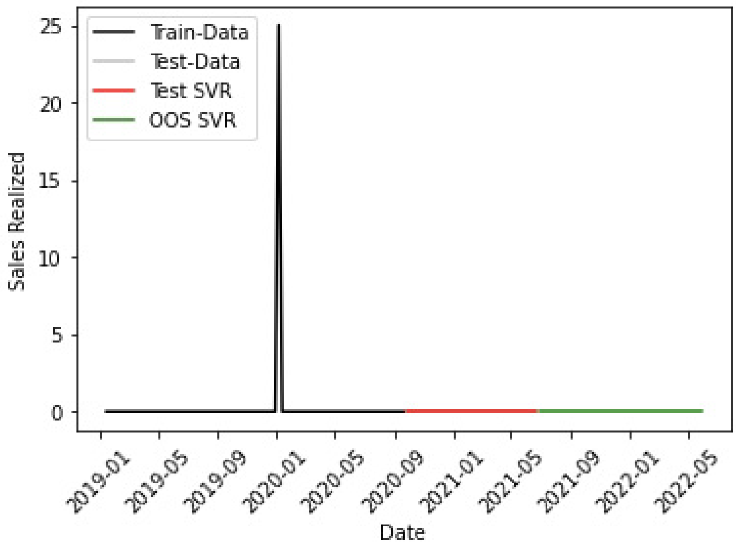


Figure 17: Plot of time series which shows 100% improvement with respect to the naive model