Silesian University of Technology
Faculty of Automatic Control, Electronics and Computer Science

Summary of Doctoral Dissertation

Solutions for selected problems of demand forecasting based on machine learning methods and domain knowledge

Joanna Badura

Supervisor: dr hab. Marek Sikora, Associate Professor

Gliwice 2023

1 Introduction

The dissertation explores the use of machine learning methods and domain knowledge to address selected demand forecasting issues.

Understanding and accurately forecasting product demand is critical for businesses and has major impacts on inventory management, supply chain efficiency, and customer satisfaction. Methods based on judgemental forecasting and expert decision-making are still used [18], but increasingly, especially in large-scale cases, methods of artificial intelligence are being developed for this task. There are various machine learning techniques available for demand forecasting. They can identify complex patterns and they are based on data. Because of this, they are helpful when there are numerous variables affecting demand.

The first thing that comes to mind when thinking about demand forecasting is the challenge of forecasting a sale of a product for which we have historical data. In this scenario, the sale of a certain product can be represented as a time series.

However, demand forecasting issues are not limited only to forecasting demand for products based on existing historical sales data for a specific product. The topic of demand forecasting is much broader, and a given dissertation devotes attention to different aspects connected with this topic. Some of the aspects addressed were identified as a research gap in the usage of machine learning methods.

In the dissertation, the author's taxonomy of problems of demand forecasting is presented. The chosen issues were then further examined in-depth. The study used actual data from companies in several industries. These are frequently collections that span several years and include various products. Most of them are proprietary. In the dissertation work, the words *sale* and *demand* are used interchangeably.

The author of the dissertation is a holder of European Unionscholarship through the European Social Fund, grant InterPOWER (POWR.03.05.00-00-Z305).

1.1 Thesis statement

By breaking down the forecasting task into specific issues and taking into account domain knowledge, useful forecast values can be obtained using machine learning methods.

Supporting thesis statement

The domain knowledge can be included by introducing new attributes or new data records from similar problems into the model, resulting in an improvement in the final forecast results.

1.2 Objectives of the work

The first aim of the dissertation was to propose a taxonomy of demand forecasting issues that can be solved by machine learning methods. The next objective was to propose the use of machine learning methods and domain knowledge for selected issues. The selection of these issues was determined by research projects carried out.

1.3 Structure of the thesis

The thesis consists of 7 chapters. Chapters 1 and 7 are respectively the introduction and summary of the thesis. Chapter 2 is an introduction to the topic of demand forecasting using machine learning and domain knowledge. Chapter 3 presents the first selected issue related to demand forecasting, namely focussing on the topic of promotional effectiveness forecasting. Chapter 4 focusses on the problem of top-down forecasting, that is, the breakdown of a higher-level forecast (e.g. for a group of products) into a lower-level forecast (e.g. a forecast for a specific product). Chapter 5 presents the next issue highlighted in the taxonomy, namely the manner and effectiveness of using additional time series to improve the forecasting of the phenomenon under investigation. Chapter 6 focusses on the issue of forecasting the probability of return of a specific product.

2 Demand forecasting – approaches and challenges

The dissertation presents the author's taxonomy of demand forecasting issues. They are divided into three categories: cases connected with short-term forecasting, medium-term forecasting, and long-term forecasting. This taxonomy is dictated by various business cases. Each issue was described in the thesis, and the problems that may arise with the issue were addressed. Figure 1 presents a novel breakdown of challenges and tasks related to demand forecasting proposed in this work. Several of the issues were selected for further research.

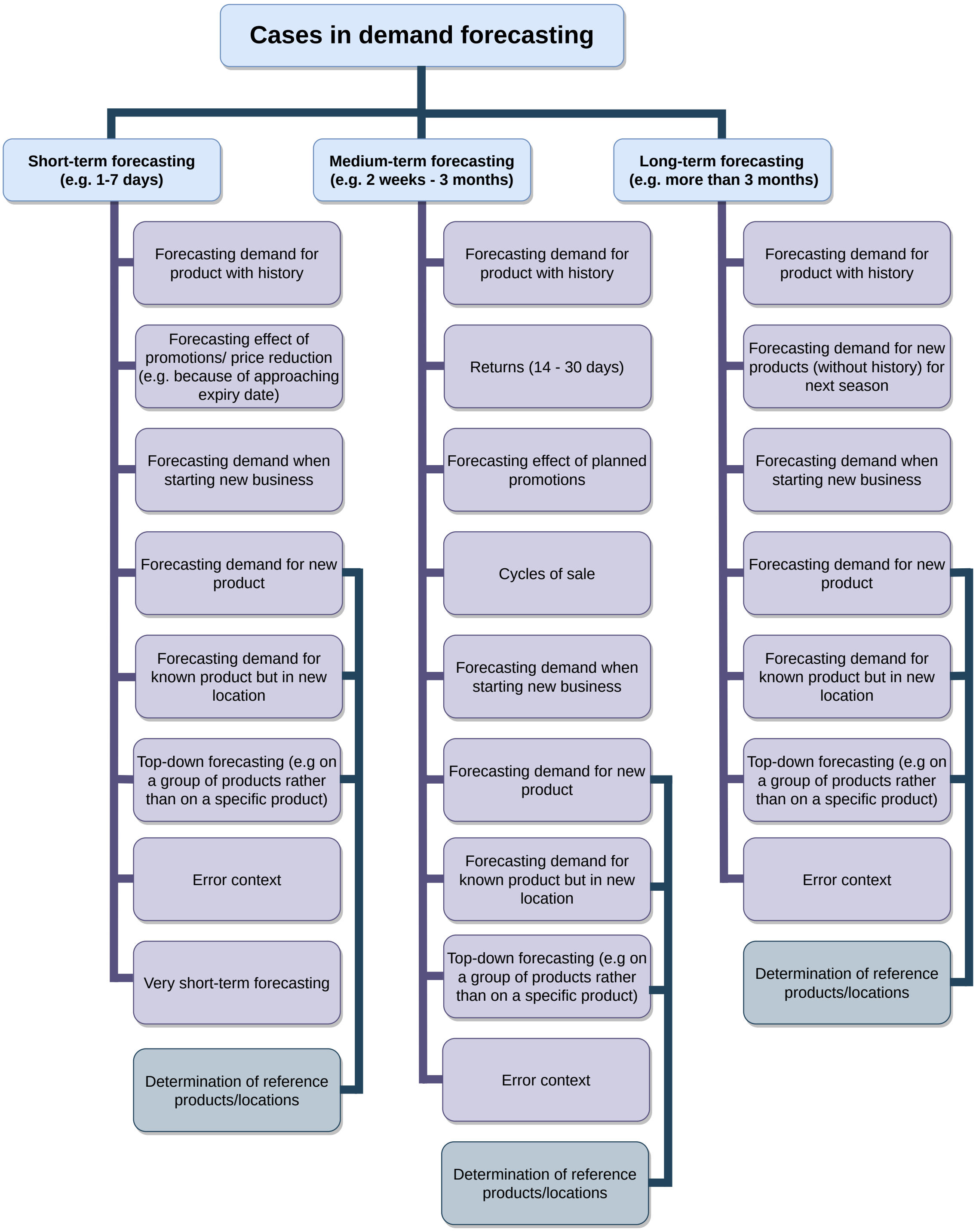

Figure 1: Breakdown of demand forecasting issues.

3 Promotions

The first topic chosen for further analysis was the subject of forecasting the effectiveness of promotions. In the first stage of the research, six indicators were proposed to accurately describe the effect of promotions [11]. Then studies were carried out on the modelling of the indicators concerned. The process of preparing the training data and feature engineering was shown in the thesis. In the first approach, indicators forecasting models were created based on historical data of all products in a group of products rather than individual products. A comparison of the XGBoost (eXtreme Gradient Boosting) [5] model, Sequential Deep Learning Models and Deep Learning Models from H2O was done. Taking into account the measure that was optimised, the best results were obtained for XGBoost in 10 out of 12 cases. The research was also presented in the paper [10].

In the next approach, the possibility of forecasting selected indicators based on product-only history data was tested [2]. Comparison of algorithms from 4 well-known families: tree-based, distance-based, neural networks, and a linear model was performed. The analysis of the correlation (R) metric values was done for each method and each product under study. This measure was relatively high, but for some products, it even dropped to 0.64. The concerning large variance of results was investigated and it turned out that the correlation R of the models correlates with the size of the historical data set that describes the promotions. This confirmed the original assumption that, for selected products, for which there is a lot of data on historical promotions, predicting the effect of a promotion based only on the data of the product in question may be possible. However, for many products, there may be too little of these data, in which case it may be useful to forecast the effect of the promotion by taking into account historical data on multiple products from the same product group.

The practical application of the resulting predictive models created for the product groups was then presented. In the first iteration of the study, the aim was to use previously prepared data and predictive models to create a prototype of price sensitivity simulator – a solution for modelling the impact of promotions on sales and profit. The user could indicate the parameters of the future promotion, and in the output, they were to receive information on how the sales of the product and the profit change for different price values. The aim of the second iteration was to propose a prototype of a promotional recommendation system. It indicated the characteristics of the best promotions, i.e. such a promotion that for the given input parameters obtains the highest forecast for the indicator describing the average number of sold units or kilogrammes each day. Among the input parameters were constraints that controlled the values of other indicators for the recommended promotion. Such a solution made it possible to find a promotion that is suitable, taking into account several indicators at the same time. The dissertation presents the interface and logic of the presented solutions.

The final section of the chapter on promotion presents the possibility of using an innovative method of Survival Action Rules induction [1] to find recommendations that can improve the effect of promotion. The dissertation presents a theory on survival analysis [21, 24], rule mining [20], action mining [6] and survival action rules. Survival action rules are a combination of typical action rules generated for classification problems and survival rules. This kind of rule indicates the action that should be made, described in the premise, to change the survival curve, which is included in the conclusion of the rule. The proposed algorithm for the induction of survival action rules is a covering rule induction algorithm and consists of two stages: growth and pruning. The details of the algorithm in question, together with pseudocodes, are included in the dissertation.

In the study conducted, the survival event was a drop in the value of the indicator that described promotion below a preset threshold. Survival time described the day of the promotion on which the event took place or, in the absence of such an event, the duration of the promotion. This representation of a given situation allowed use of the survival action rule induction method. The resulting set of rules showed changes (actions) in the promotion parameters that would reduce the probability that the indicator fell below a given threshold during the promotion. All 10 rules obtained had a log-rank test p-value of less than 0.01, which indicated that they were of good quality.

The case study of using survival action rules for promotions was presented in the thesis on the example of one selected rule. The premise of this rule is described by Table 1 and the survival curves for the source and target rule are shown in Figure 2. The premise of the rule consists of five conditions, including four actions and one elementary condition (stable attribute). Actions were inducted for attributes: number of all promotions in the store that were advertised on TV, radio or internet (*nb_promos_TRI*), number of all promotions that were advertised in a different way (*nb_promos_D*), *price* and number of all promotions in a store (*nb_promos*). The premise of the rule shows what changes should occur in the values of these attributes to change the survival curve for the indicator under study (presented in Figure 2). The Survival Action Rule induction algorithm and the presented case study showed a new way of recommending changes that could be applied to make the promotion meet the expected value of the indicator for longer time.

Premise
(nb_promos_TRI < 18 → nb_promos_TRI > 2) AND
(nb_promos_D < 118 → nb_promos_D > 22) AND
(price < 1.57 → price > 1.64) AND (first_date_day_of_year < 339) AND
(nb_promos ∈ <37, 302) → nb_promos > 143)

Table 1: Premise of analysed survival action rule.

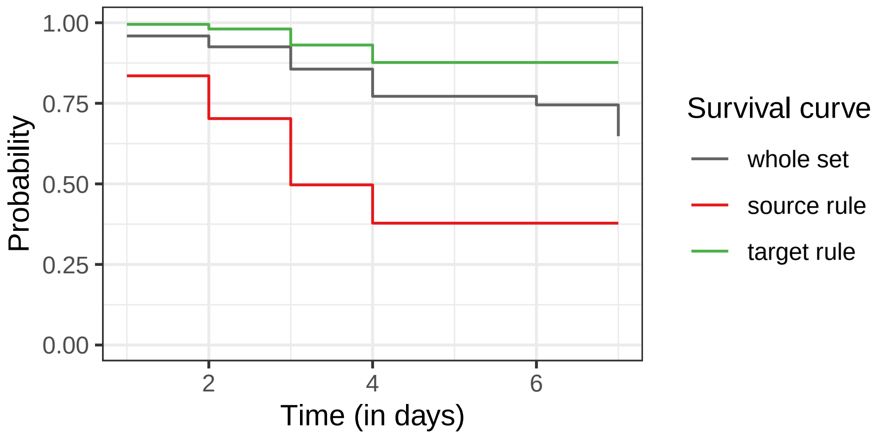

Figure 2: Survival curves of rule conclusions for rule which premise is presented in Table 1. The curve corresponding to the source rule is in red, the curve corresponding to the target rule is in green.

4 Top-down forecasting

Another study looked at the problem of top-down forecasting, that is, the problem of splitting higher-level forecasts into lower-level forecasts [25]. A higher-level forecast can be a forecast of a demand for a whole group of products suggested by domain expert or it can also be a forecast derived from another predictive model. A lower-level forecast is a forecast of demand for a particular product. In general, the task of top-down forecasting was to find the weight w by which the forecast for the entire category should be multiplied to get the forecast for the particular product.

In the research, the naive method, the custom nearest neighbour approach, the parametric linear mixed model, and the XGBoost method were tested. Studies were carried out based on a data set from a clothing retailer. The forecasts for a certain product category and product-specific forecasts were provided to us by a business partner. These product-specific forecasts were considered as a baseline for the results obtained. The research presented in this chapter was also partially presented in the paper [12].

The naive method calculated the weight w based on sales exactly one year earlier.

Next, the K-Nearest Neighbours (KNN) [16] method was chosen as an extension of the naive method. The KNN algorithm made a weight w forecast by searching for the most similar observations from the entire range of historical data. The thesis presented a special preparation of the input data. An important approach concerned changing the values of the *size* attribute, which had a different convention for different products. To calculate the distances between different observations (different products) using the size attribute, the historical dataset was divided by size – a separate subset for each category of products, which had different clothing sizes. Then, for each of the sizes in the set, the percentile values within those subsets were determined, and it was a new value for the attribute. The nominal attribute *product type* also received additional entry data. The *product types dissimilarity table* was proposed in order to obtain the distance between different types of products.

Another tested model was Linear Mixed Models (LMM) which is an extension of simple linear regression models and can be used for data with a hierarchical structure, which is observed in fashion. These models incorporate fixed and random effects. In this approach, a model for each product group was estimated with random effects defined by product type, size, and colour. Linear mixed models also used *product types dissimilarity table* and changes in attribute values similar to the KNN approach.

The last method examined was the XGBoost method. The model was also tasked with forecasting the weight by which the category forecast should be multiplied to obtain a product-specific forecast. The forecasting model used the same features as the KNN and LMM models.

Statistical tests were used to evaluate the differences of the prediction errors between the tested methods. First, a nonparametric Friedman test [8] was conducted, which is used to compare the results of more than two algorithms with each other. This test checks the null hypothesis that all algorithms are equal. The null hypothesis could be rejected, so a *post-hoc* analysis was performed using the Nemenyi test [19] and the results were presented using Critical Difference diagram (CD diagram) [7]. This diagram shows groups of algorithms that are significantly different from each other on the basis of the Nemenyi procedure.

The ranking of the methods indicated that considering all the products, it was the LMM method without optimisation that gave the best results. The CD diagram did not show that there was a statistically significant difference between the average ranks of the LMM method without optimisation and the baseline; however, the Nemenyi test is considered conservative, so the Wilcoxon test with correction was also performed [22]. This test yielded significant differences at the 0.05 level, and it means that based on the Wilcoxon test we can conclude that the proposed LMM method without optimisation gave significantly better results than baseline.

The research carried out proposed the use of machine learning methods to perform the top-down forecasting task. In the context of forecasting for the fashion industry, it seems worth noting the method used here to prepare the data in such a way that it can be used in methods that use a distance measure. The attribute *size* was changed to a numerical value using percentile values that correspond to a particular size in the historical data set. For *product type* attribute, which served as identifiers and had no further significance, the *product types dissimilarity table* was proposed in order to obtain distances between analysed products. Such a table can be determined from data (based on information on more general product categories), but can also be determined by an expert.

5 Incorporating additional time series related to demand forecasting

The next chapter presented the possibilities and advantages of using additional time series to improve the forecast performance of the issue under consideration.

5.1 Incorporating data from similar locations

First, the possibility of using analogous time series from other locations was shown. Different possibilities were presented to group and add information from similar time series. The following approaches were tested:

1. Univariate forecast – Forecasting models were created based only on historical data from the location under consideration. No additional time series were used. The results of these experiments were our benchmarks. The methods that were taken into account were: moving average, ARIMA [3], Prophet [23], XGBoost, and Long Short-Term Memory Network (LSTM) [14].

2. Forecast after putting all datasets into one group – In these experiments, data from all locations were used in the forecasting model. Time series were not grouped in any detailed way. The LSTM method and XGBoost in two versions were considered.

3. Forecast based on similar records – In this experiment, a typical KNN regression method was used.

4. Forecast based on clustering time series – The datasets were grouped by the nature of their normalised time series describing decision value. Then, on the basis of these groups, forecasting models were created. The LSTM method and XGBoost in two versions were considered.

5. Forecast based on double clustering – In this experiment data from all time series were normalised and grouped by the characteristic of the time series. A column describing the membership of the designated group was then added to each normalised set. Then, all datasets were combined into one dataset. Next, an optimal number of clusters was calculated for algorithm k-means, and then k-means clustering was performed on the created dataset. For each group, the XGBoost model was created.

Each of the approaches is described in detail in the dissertation.

Two data sets were used in the study. The first set described the sale of petrol at many petrol stations and was used in the research presented in [9]. The second described the usage of liquefied petroleum gas (LPG) in tanks located near buildings [17]. These two groups of datasets contained only a few additional conditional attributes, so feature engineering was a crucial step.

Because many approaches were tested on multiple data sets, the results for the Friedman test and Iman Davenport's correction of Friedman's rank sum test were obtained. Then, *post-hoc* analysis – the Nemenyi test and Wilcoxon test with p-value correction for multiple comparisons – were carried out. Taking into account CD diagrams presenting results of Nemenyi test and analysing the results of Wilcoxon tests, the best results were always obtained for methods that used data from other locations. In both cases, in the best group indicated by CD diagrams, the best approaches included the same 3 procedures. The first one was a proposed double clustering approach. The second method involved combining all datasets into one train dataset, which was used to train a model that predicted the normalised value of a decision value. The third approach used time series clustering. In this approach, after creating groups based on the characteristics of the time series, a separate training set was created for each group. Then, the data describing the time series allocated in one group were combined (after being normalised), and then predictive models were trained on the basis of these sets, which predicted the normalised value within these groups. For LPG usage predictions also ARIMA performed significantly well. Considering the given results, it can be concluded that for the cases studied, the introduction of data on a similar problem from other locations significantly improved the prediction results and the given approach may be worth considering in similar situations.

5.2 Using data from the same location

Another issue similar to the problem of incorporating data from similar locations is incorporating data from the same location. The focus was on adding data from different time series that can be measured in the same location as a predicted time series in order to obtain better forecasts. This issue was illustrated by the example of forecasting energy consumption in households. The research presented was carried out in the context of a project that deals with the development of a digital twin model of a building. Thanks to this, it was possible to use information about the energy consumption of selected household equipment. These data, with information about the total energy consumption of the location and the weather data, were used in forecasting approaches. As far as the author knows, this approach is unique in comparison with other work. The research was also presented in the paper [13].

Five different locations, equipped with devices with energy consumption metres at each site, were examined. Energy consumption for the entire location was also monitored. In the experiments, the baseline model was established using a naive approach. This model predicted the value observed at a given location exactly one week earlier. Then four models based on linear regression were created. LSTM and Prophet methods were employed as more advanced methods for time series forecasting tasks. To assess the improvement of forecasts with different attribute sets, we considered the following scenarios:

- Using information only about total energy consumption.
- Using information about total energy consumption and weather.
- Using information about total energy consumption and energy consumption of the top 10 energy-consuming devices.
- Using information about total energy consumption, weather, and energy consumption of the top 10 energy-consuming devices.

The goal of the studies was to obtain less than 25% error of energy consumption prediction and it was achieved for 4 out of 5 locations. For these four locations, the best results were obtained in the experiment, where the Prophet method was used with information on the total energy consumption and the energy consumption of the top 10 energy-consuming devices.

The dissertation also presents a method for analysing the errors of the forecasting model. For this, classification model was created for each location. For each location and its Prophet model, a prediction was labelled as *Correct* when the error for the prediction was below 25% and *Incorrect* when the error for the prediction was equal to or greater than 25%, or when the model failed to produce a prediction. These were treated as two labels in the classification task. Then, decision tree models were created using custom features. These features described the energy consumption compared to previous historical values and the trend change in energy consumption. A detailed description of feature engeenering process for creating decision tree model was included in the dissertation. The decision models revealed which features (utilisation of specific devices) had the greatest impact on the performance of the forecast. Last but not least, created decision tree models were used to limit the number of monitored devices. In most cases, when removing information about devices that appeared in the tree from the Prophet model, we obtained results very similar to those of when all 10 devices were monitored.

Summarising the research presented in this chapter of dissertation, the potential of incorporating information from other time series, originating from either the same or different locations was analysed. In all the cases examined, additional information led to better forecasts.

6 Returns

The final issue considered was the problem of predicting the probability of a return of a particular piece of product from the purchase transaction. The possibility of using machine learning methods for this task was checked. A review of the literature indicated that the topic of forecasting returns using machine learning methods has not been widely analysed, and it may be a possible research gap.

The idea was to create a predictive model based on data that describes purchase transactions. For each product in each transaction, it was intended to add a decision binary attribute that would indicate if a product was returned. However, after analysing the available data it occurred that in the retailer database, there was no information on which return transaction was connected to which purchase transaction. For this reason, there was no way to simply receive information about whether the product was returned. Due to this fact, a custom algorithm for linking sale transactions with return transactions was proposed.

The algorithm consists of two steps. In the first step, for the return under consideration, an attempt is made to find a purchase transaction that has a high probability of being a corresponding purchase transaction. This will be a purchase transaction that took place within a limited period of time before the return and contains all of the products that were included in the return. If there is only one such transaction in the database, it can be said that the corresponding purchase transaction has been found, and the algorithm is terminated. If there are many purchase transactions that meet the criteria, the algorithm proceeds to Step 2. If no purchase transaction meets the conditions of Step 1, then the algorithm is also terminated with the information that no matching purchase transaction was found for the return under consideration.

In Step 2, an attempt is made to select a matching purchase transaction among the previously narrowed group. Selection of a matching purchase transaction is based on the similarity of the transaction channel (the purchase and return took place online or in a store), on the number of similar products in the purchase transaction (many similar items may suggest that the customer bought products with the intention of returning some of them) and based on the value of the shopping cart. If after all these steps there is still more than one matching purchase transaction, the final selection of the matching purchase transaction is done randomly but already in a heavily restricted group. Block diagrams of the proposed algorithm steps are presented in the dissertation.

Due to the lack of ground truth on the links between transactions, it is not possible to verify the proposed algorithm. However, it can be noted that many return transactions were associated with high confidence with one purchase transaction (40.6%). For the remaining transactions, only a few of them did not find any matching purchase transaction (2.9%) and a small number of returns were linked using random selection (1.1%). Taking this into account, the proposed algorithm appears to be a good solution to combine return transactions with purchases in cases where there are no clear relationships. This link allows for the creation of a decision attribute in the training set and the creation of other conditional attributes that describe the purchase transaction.

In the research presented, three algorithms were taken into account for training decision model that would predict whether the product would be returned: logistic regression [15], Random Forest [4] and XGBoost. All algorithms were tested in 4 configurations.

In the first approach, only the original attributes that describe each product were used. These included colour, size, category of product, and price. The dataset also described the date of the product sale, including information about holidays and special events.

The second approach extended the former. The feature engineering was performed and the attributes describing the basket and other bought products were added. This is based on domain knowledge, conclusions from the literature review, and our own observations of the data – when multiple pieces of very similar products are purchased, or the same ones only in different sizes/colours, the probability of a return on a given transaction increases. The new attributes were, for example, number of products in the basket, the value of the basket, number of products in the basket in the same colour but different sizes and the popularity score.

In the third approach, the tuning of parameters for each model was performed. The cut-off value was also optimised. In this approach, all previously created attributes were used.

The last approach included class balancing. The Synthetic Minority Over-sampling Technique (SMOTE) was used. This approach included all the previous modifications.

The best result was obtained for XGboost for Approach number 3 (after feature engineering and parameter tuning) with the value of AUC equal to 0.806. This is a satisfactory result. Another conclusion is that the feature engineering and parameter tuning performed gave good results. It can be concluded that the addition of features describing the content of the shopping basket improved the results for each method. These features were able to be added thanks to the algorithm created, linking the purchase transaction to the sale transaction. Although we did not have real information on the connections between the transactions, it was possible to obtain features that improved the predictive model. This may indicate the usefulness of the proposed algorithm.

7 Summary

In the thesis, the original taxonomy of demand forecasting issues was presented. Each task was described taking into account specific problems that may be associated with each of them. This taxonomy can show a broader picture of the issues involved in demand forecasting. Then, attention was turned to selected issues and consideration was given to how machine learning methods can be applied to solve given problems. The thesis does not focus on the typical issue of time series forecasting where one predicts future data points based on historical demand data. The objective was also not to compare all known time series forecasting algorithms. Instead, the dissertation concentrates on lesser-known demand forecasting issues.

The contribution of the works is as follows:

- The taxonomy of demand forecasting issues was proposed, allowing for a structured approach to the topic, and providing a broader perspective than only focusing on time series forecasting.
- A way of simulating the effect of promotion using six indicators and a machine learning models was proposed. This approach was then used for price sensitivity simulator and recommendation system for promotions.

- An original algorithm for Survival Action Rules induction was presented to indicate changes that would affect the effectiveness of the promotion.

- The use of machine learning methods for top-down forecasting in fashion sector was presented. A method of feature preparation was proposed using *product types dissimilarity table* and the conversion of different product sizes into a numerical characteristic.

- Studies were conducted that indicate the importance of using additional time series to forecast demand for the scenarios considered, taking into account time series from similar locations and the same location.

- Using information from the digital twin model of a building, additional time series were used for creating predictive model. Our research has shown that monitoring energy consumption not only for an entire location, but also for selected appliances, can significantly improve the ability to predict future energy usage.

- An algorithm was proposed to combine the return transaction with the purchase transaction. The topic of forecasting the return of a specific product from a specific purchase transaction was also considered, achieving good forecasting results.

- The contribution also includes a presentation of the features that are worth using for specific issues in the forecasting models and the process of feature engineering.

- The dissertation presented lesser known issues related to demand forecasting.

The research conducted also confirmed the thesis statement. The proposed taxonomy of demand forecasting problems allowed them to be approached in custom way using domain knowledge and machine learning methods. Domain knowledge was introduced in the research by making a decision about the range of data used in the predictive model, in the context of what is to be forecast and for which forecast horizons. In all studies, domain knowledge also manifested itself in the context of feature engineering.

All research was conducted on real data, mostly sourced from large companies with multiple branches. Data sets also often spanned several years, came from multiple locations, and involved multiple products. The studies considered were dictated by R&D research projects, so it can be concluded that these are relevant to the business. Part of the research presented in the dissertation was linked to the following projects: "Decision Support and Knowledge Management System for the Retail Trade Industry (SensAI)" (POIR.01.01.01-00-0871/17-00) and "The SupplAI system developed for the supply chain automatization and optimization utilizes artificial intelligence technics – including sales forecast – to gain control over the big data online analysis, aiming for efficient cost reduction managementand supporting and improving the retail governance sector" (POIR.01.01.01-00-0500/21) – both projects were co-financed by European Funds and realised with the company 3Soft S.A. "System for decision support and management of operational and process knowledge for the LPG gas distributors market" co-funded by the Polish agency National Center for Research and Development within grant POIR.01.01.01-00-0104/17. Research related to forecasting fuel sales at gas stations was realized in co-operation with FuelPrime Department in AIUT Ltd. and was partially supported by the European Union through the European Social Fund (grant POWR.03.05.00-00-Z305).

Files with data for the short-term LPG demand forecast task are openly available in `https://github.com/adaa-polsl/dss4lpg`. The website also includes software that allows to perform simple analyses and to generate a summary report. Data sets describing the time series of fuel sales were made available on the website `http://adaa.polsl.pl/index.php/datasets-software/`.

References

[1] Badura, J., Hermansa, M., Kozielski, M., Sikora, M., Wróbel, Ł. Separate-and-conquer Survival Action Rule Learning. *Knowledge-Based Systems*, 2023 (Submitted to the journal Knowledge-Based System. Article after first round of peer review. Status as of 22/05/2023).

[2] Blachnik, M., Henzel, J. Estimating the Performance Indicators of Promotion Efficiency in FMCG Retail, *Neural Information Processing*, pages 320–332. Springer International Publishing, 2020.

[3] Box, G. E., Jenkins, G. M., Reinsel, G. C., Ljung, G. M. *Time series analysis: forecasting and control.* John Wiley & Sons, 2015.

[4] Breiman, L. Random Forests. *Machine Learning*, 45(1):5–32, 2001.

[5] Chen, T., Guestrin, C. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, volume KDD '16, pages 785–794. ACM, 2016.

[6] Dardzinska, A. *Action rules mining*, volume 468. Springer, 2012.

[7] Demšar, J. Statistical Comparisons of Classifiers over Multiple Data Sets. *J. Mach. Learn. Res.*, 7:1–30, 2006.

[8] Friedman, M. The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association*, 32(200):675–701, 1937.

[9] Henzel, J., Bularz, J., Sikora, M. Impact of time series clustering on fuel sales prediction results, *Position and Communication Papers of the 16th Conference on Computer Science and Intelligence Systems*, volume 26 of series *Annals of Computer Science and Information Systems*, page 13–21. PTI, 2021.

[10] Henzel, J., Sikora, M. Gradient Boosting and Deep Learning Models Approach to Forecasting Promotions Efficiency in FMCG Retail, *Artificial Intelligence and Soft Computing*, pages 336–345, Cham, 2020. Springer International Publishing.

[11] Henzel, J., Sikora, M. Gradient Boosting Application in Forecasting of Performance Indicators Values for Measuring the Efficiency of Promotions in FMCG Retail, *Proceedings of the 2020 Federated Conference on Computer Science and Information Systems*, volume 21 of series *Annals of Computer Science and Information Systems*, pages 59–68. IEEE, 2020.

[12] Henzel, J., Wawrowski, Ł., Kubina, A., Sikora, M., Wróbel, Ł. Demand forecasting in the fashion business — an example of customized nearest neighbour and linear mixed model approaches. *Proceedings of the 17th Conference on Computer Science and Intelligence Systems (FedCSIS)*, pages 61–65, 2022.

[13] Henzel, J., Wróbel, Ł., Fice, M., Sikora, M. Energy Consumption Forecasting for the Digital-Twin Model of the Building. *Energies*, 15(12), 2022.

[14] Hochreiter, S., Schmidhuber, J. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 1997.

[15] Hosmer, D. W., Lemeshow, S. *Applied Logistic Regression*. Wiley, 2nd, 2000.

[16] Kordos, M., Strzempa, D., Blachnik, M. Do We Need Whatever More than k-NN? *LNCS*, 6113:414–421, 2010.

[17] Kozielski, M., Henzel, J., Wróbel, Ł., Łaskarzewski, Z., Sikora, M. A Sensor Data-Driven Decision Support System for Liquefied Petroleum Gas Suppliers. *Applied Sciences*, 11(8), 2021.

[18] Lawrence, M., Goodwin, P., O'Connor, M., Önkal, D. Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, 22(3):493–518, 2006.

[19] Nemenyi, P. B. *Distribution-Free Multiple Comparisons*. PhD thesis, Princeton University, 1963.

[20] Sikora, M. Wybrane metody oceny i przycinania reguł decyzyjnych. *Studia Informatica*, 33(3B):5–331, 2012.

[21] Stevenson, M. An Introduction to Survival Analysis. EpiCentre, IVABS, Massey University, 2007.

[22] Stąpor, K., Ksieniewicz, P., García, S., Woźniak, M. How to design the fair experimental classifier evaluation. *Appl. Soft Comput.*, 104:107219, 2021.

[23] Taylor, S. J., Letham, B. Forecasting at Scale. *The American Statistician*, 72(1):37–45, 2018.

[24] Wróbel, Ł., Gudyś, A., Sikora, M. Learning rule sets from survival data. *BMC Bioinformatics*, 18(285):1–13, 2017.

[25] Zotteri, G., Kalchschmidt, M., Caniato, F. The impact of aggregation level on forecasting performance. *International Journal of Production Economics*, 93-94:479–491, 2005.