BABEŞ-BOLYAI UNIVERSITY CLUJ-NAPOCA

FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

PHD THESIS

Model Explanation Methods for Decision Support in Demand Forecasting

Mátyás Kuti-Kreszács

Scientific supervisor: Prof. Dr. Dioșan Laura

2025

Keywords: demand forecasting, explainable AI, decision support, machine learning, time series forecasting

I hereby declare that I am the sole author of this thesis.

I authorize Babeş-Bolyai University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Mátyás Kuti-Kreszács

I further authorize Babeş-Bolyai University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Mátyás Kuti-Kreszács

Acknowledgements

TODO: include

Dedicated to ..

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction

Product demand forecasting is a common business problem in many industries. But especially in production planning, manufacturing, logistics, inventory management, retail, and marketing.

1.2 Objectives

The motivation of the thesis comes from the need to improve the explainability of machine learning forecasting models in order to raise trust in the system and deliver actionable insights for decision support.

1.3 Contributions

- Contribution 1: Challenges in explainable AI for demand forecasting
- Contribution 2: Hierarchical feature importance in demand forecasting
- Contribution 3: Delivering actionable insights for decision support

1.4 Thesis Structure

TODO: describe the structure in the end

1.5 List of Publications

1.6 Thesis Structure

In Chapter 2, we introduced the background of demand forecasting, including statistical and machine learning models, forecasting techniques, and multi-step forecasting.

In Chapter 3, we discussed the importance of interpretability and explainability in machine learning models, including the difference between the two concepts, the importance of model transparency, and the different methods to interpret and explain machine learning models.

In Chapter 4, we discussed the importance of feature importance estimation in machine learning models, the challenges in practice, and the different techniques to estimate feature importance and address these challenges.

In Chapter 5, we discussed the importance of hierarchical feature importance in machine learning models, the construction of hierarchies, and the experiments on hierarchical and grouped data.

In Chapter 6, we discussed the importance of collaborative development for decision support, the automated delivery of models as services, and the conclusions.

In Chapter 7, we discussed the conclusions and future work.

Chapter 2

Demand Forecasting

2.1 Business problem

2.2 Statistical Models

AR, ARIMA, SARIMAX models are widely used when the series is stationary and data is univariate.

2.3 Machine Learning Models

2.3.1 Demand Forecasting with machine learning

Demand forecasting is a prediction problem that aims to estimate future needs based on historical data. Statistical forecasting methods such as ARIMA(6; 13) and exponential smoothing (13) have been widely used in demand forecasting. However, they have limitations in intermittent multi-series and hierarchical forecasting, where machine learning models have shown better performance(25). An important aspect also is that there may be multiple exogenous variables so-called demand drivers(27) that can influence the demand. Internal factors such as price, promotions, and external factors like weather, holidays, and economic indicators can be considered as demand drivers. These can be used as features in machine learning models to improve forecast accuracy.

Machine learning models such as tree ensembles and neural networks have been successfully applied to demand forecasting tasks(25). Ensemble models in general can be homogeneous with individual models of the same type or heterogeneous with models of different types. We considered only homogeneous ensemble tree models because of the applicability of some model-specific explanation methods. To build tree ensembles, bagging methods such as random forest(16) can be used, which trains multiple decision trees on different subsets of the data, and the final prediction is the average of the predictions of the individual models. In addition, boosting methods such as Gradient Boosting Machines (GBM) (14), XGBoost (26), and LightGBM (10), which train models sequentially on the residuals of the previous model, in this case using the sum of individual predictions. In a notable forecasting competition (18), a LightGBM model was the winner and secured four of the top five positions.

2.4 Forecasting Techniques

- Single-step forecasting
- Multi-step forecasting

Forecasting models

- Univariate
- Independent multi-series
- Dependent multi-series or multi-variate

Forecasting techniques can be divided into single-series or multi-series forecasting from the perspective of the model's input. Single-series forecasting refers to the prediction of a single time series, while multiseries forecasting involves the prediction of multiple time series, with the same global model(22). These series can be related to each other, such as sales of different products, or they can be independent, such as sales in different regions; therefore, it is important to consider the hierarchical structure of the data.

Hierarchical forecasting refers to the prediction of multiple time series that are related to each other in a hierarchical structure(11). It can be tackled with different single-level approaches, such as bottom-up, top-down, or middle-out(11). The top-down approach would involve a single series model for the total demand and then disaggregating it to the lower levels. The middle-out and bottom-up approach would involve a multiseries model. Grouped time-series forecasting is a special case of hierarchical forecasting, where the series are aggregated based on attributes such as product type, region, or sales channel.

(27) suggests three major hierarchies in demand forecasting: product hierarchy, geographical hierarchy, and time hierarchy. The product hierarchy refers to the categorisation of products according to their attributes, such as product type, brand, or category. The geographic hierarchy involves the division of sales regions based on geographic attributes, such as country, state, or city down to the point of sale. Time hierarchy refers to the temporal structure of the data, such as year, month, week, day, and hour.

2.5 Multi-step Forecasting

- Direct forecasting
- Recursive forecasting
- Multiple-output forecasting

Chapter 3

Interpretability and explainability

3.1 Interpretability, Explainability, Transparency, and Trust

Interpretability and explainability are two important concepts in the field of machine learning and artificial intelligence.

3.2 Scope of Explainability in Forecasting

3.2.1 Feature Importance

3.3 Challenges in Explainability for Time Series Forecasting

3.4 Feature Importance as a Basis for Model Reasoning

3.5 Local explainability

3.6 Cohort explainability

3.7 Global explainability

3.7.1 Feature importance

Feature importance (FI) or feature attribution is considered an interpretation method resulting in a summary statistic that assigns a score to each input feature (19). Depending on their scope, the FI methods can be global or local (9; 19). The global feature importance (GFI) or model feature attribution methods explain the contribution of features to overall predictions, while the local FI quantifies feature contributions to specific predictions (19). Although related, GFI methods differ from feature selection, which identifies irrelevant features before training. GFI methods can be model-specific, which are limited to specific model types, while model-agnostic ones are applicable independent of the model type(19). Another categorisation of FI methods is given by how it is calculated, in which case the importance can be based on the model's structure, while the other approach relies on a dataset.

Among the model-agnostic methods, one of the most common is permutation feature importance (PFI) which was proposed to measure FI in random forests(3). It is a model-agnostic, data-dependent method that measures the decrease in the model's performance when the features are permuted. The PFI can be calculated using different metrics such as the mean squared error (MSE), the mean absolute error (MAE), or the coefficient of determination (R^2). PFI also has limitations, as it is sensitive to over- and underfitting(20), in which case the FI differs on training and test data, so the use of both datasets can be beneficial. In addition, another flaw of the PFI method is that it can generate cases in which the model does not have training data(21; 7), but other methods were proposed to overcome this(12; 15).

SHAP(SHapley Additive exPlanation)(17) values contribute local explanation for individual predictions, but aggregates of it are useful to assess the importance of global features. For example, the mean absolute SHAP values quantify the importance of the feature regardless of the direction of the impact on the prediction. There are different algorithms for approximation from which Kernel SHAP(17) is one that is model-agnostic. TShap (28) is a method for estimating SHAP values for time series data, but it uses a surrogate model, so it gives the FI of the surrogate. Another related method is SAGE *(Shapley additive global importance)* (12), which estimates the contribution of each feature to the model's performance.

Tree specific GFI methods are gain-based importance values which were already introduced with decision trees (8) It measures of the reduction in mean average error(MAE) made by the decisions based on the respective feature. Another measure is the split-based importance(26) refers to the number of decisions made by the model based on a feature. The previously presented SHAP also has a tree model-specific solution for approximation, called TreeSHAP (24)

3.8 Explainable AI in Forecasting

3.8.1 Explainability in forecasting

The number of publications on forecasting explainability is limited. (5) tackled the presentation of explanations for sales forecasting models, but not the explanation methods themselves. (23) used SHAP values to explain the prediction of a time series model but on local level and not global level. Skforecast citejoachim2023demand library extracts model specific global feature importance from tree ensemble models. The work is focused on either global feature importance or local feature contribution without considering the multi-series and hierarchical structure of the data.

3.8.2 Feature importance as a basis for model reasoning

Feature importance methods can provide insight into the model's decision-making process and help to understand the underlying rules and reasoning behind the predictions. By including demand drivers as features in the model, the feature importance methods can help to identify the key drivers of demand. For external factors such as weather, holidays, and economic indicators, the importance of the characteristics can help to understand their impact on demand. Through internal factors like price, promotions, the feature importance can help to understand post-promotion effects and the impact of price changes on the demand(27). Knowing the influence of internal factors can help to optimize pricing strategies and promotional activities. However, causation and correlation are different concepts, and the feature importance methods can only provide correlation; therefore, the identified key features should be further analyzed to understand the causation(3)

3.9 Conclusions

Chapter 4

Feature Importance Estimation in ML Forecasting models

4.1 Techniques

4.2 Challenges in practice

4.3 Dataset simulation

4.4 Model training and evaluation

4.5 Addressing feature importance estimation challenges

4.6 Conclusions

Chapter 5

Hierarchical feature importance

5.1 Interpretability and explainability

5.2 Hierarchy construction

5.3 Experiments on hierarchical and grouped data

Our research design5.1 includes the following steps:

- Data collection: identify datasets with hierarchical and grouped time series data describing sales/demand for multiple product categories and regions with exogenous variables.
- Tool evaluation: assess the applicability of existing libraries for hierarchical forecasting and XAI techniques.
- Model implementation: we build global models that consider multiple series and exogenous variables.
- Feature importance analysis: We apply model attribution methods and aggregation and decomposition techniques to identify key features and analyze their impact on the forecast.
- Model reasoning: analyze the feature contributions to forecast and identify underlying rules on different levels of the hierarchy.

5.3.1 Dataset

To model the hierarchical impact of features on forecasting, we must use datasets with multiple series and exogenous variables that represent demand drivers. There are multiple open sales data sets available; however, there are just a few, such as the M5 competition (18) and the Kaggle datasets (4) that include exogenous variables. For our initial exploration, we sampled M5 competition (18) dataset, which includes sales data for multiple product categories and regions. The dataset contains daily

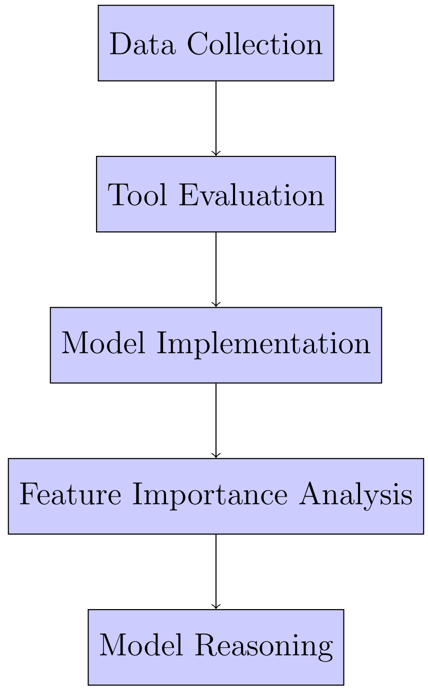

Figure 5.1: Research design

sales information for 3049 products in 10 stores over 5 years. For our analysis, we identified three products that have similar sales patterns and are sold in two states and five stores. As products are from the same category and department, the hierarchy at the product level was not considered. The reason for this filtering is to reduce the complexity of the model and to focus on the feature importance analysis. The selected products are FOODS_3_586, FOODS_3_080, and FOODS_3_555 and are sold in three states of Texas (TX), Wisconsin (WI). The total sales data for these products are shown in Figure 5.2. Our hierarchical structure is shown in Figure 5.3.

5.3.2 Model implementation

The modelling approach is to build a single global on all series and exogenous variables for bottom-up aggregation. For creating forecast models, the skforecast(2) library was used. The base model for hierarchical forecasting was LightGBM (10) due to its efficiency and also because of its widespread usage in the M5 competition in this

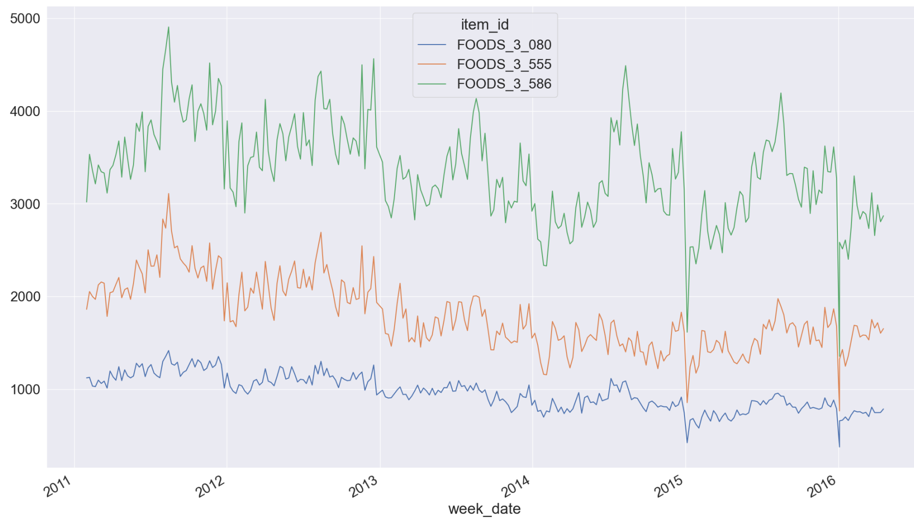

Figure 5.2: Total weekly sales for the chosen products

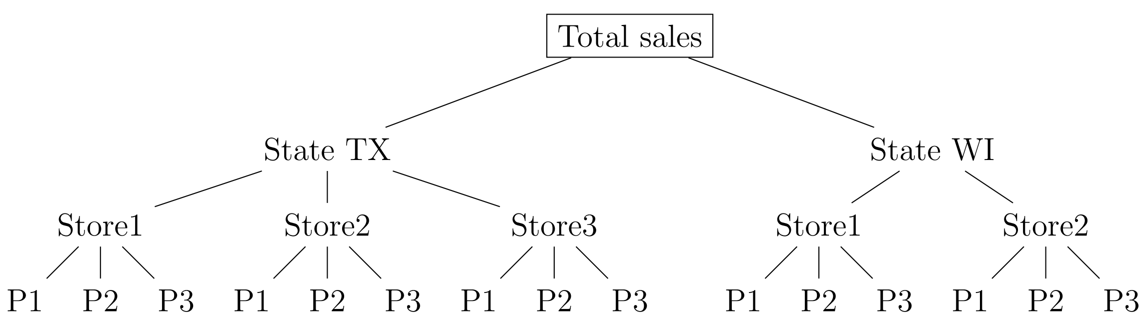

Figure 5.3: Hierarchical structure of product sales data (P1-3 = Product1-3)

Parameter	Search space	Description
n_estimators	50-1000	Number of boosting iterations
max_depth	5-50	Maximum depth of the tree
min_samples_leaf	1-10	Minimum number of samples required to be at a leaf node
num_lagged_sales	4-52	Number of lagged sales records used as features

Table 5.1: Model hyperparameters search space

data set(18). Other ensemble models such as Random Forest or Gradient Boosting Machines could be used as well. Other reasons for choosing LightGBM are that it can handle categorical variables without the need for one-hot encoding, and that it supports model-specific split and gain-based global feature importance methods.

Hyperparameter tuning was performed using the Optuna library(1), by Bayesian optimisation. The search space5.1 was defined for the parameters of the LightGBM model, including the number of predictors, the minimum number of samples in the leaf, and the maximum depth of the tree. In addition, the number of lagged sales records used as features was included in the search space. For the search, the data was split into training and validation sets, the last year being the validation set used for backtesting. The performance of the model was evaluated as a mean square error (MSE) in the validation set for each configuration. The best configuration found was with 239 estimators and a maximum depth of 26 with a backtesting MSE 4263.01 The lagged sales records used as features were 1, 4, 5, 13, and 52 weeks.

The feature input for the final model is a table with the following columns:

- week_of_year represented as numerical values (1-52)
- sell_price for the week for the product in the store
- num_of_events for the week
- snap_days for the week in the state
- lag_n for n in [1, 4, 5, 13, 52] representing the sales from the previous weeks

- series_id noted as (_level_skforecast) encoded as a numerical value representing the series hierarchy

5.4 Conclusions

Main findings

Chapter 6

Collaborative development for decision support

6.1 Best Practices for ML System Development

6.2 Automated delivery of models as services

Best Practices for Collaborative Development of ML Systems

- Insights from the development of a demand forecasting system.
- Discussion on deploying scalable and interpretable ML models in industrial environments.

6.3 Conclusions

Chapter 7

Conclusions and Future Work

7.1 Summary of Contributions

7.2 Future Research Directions

Bibliography

[1] Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [17]

[2] Amat Rodrigo, J. and Escobar Ortiz, J. (2024). skforecast. [15]

[3] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32. [9, 10]

[4] Corporación Favorita inversion, Julia Elliott, M. M. (2017). Corporación favorita grocery sales forecasting. [14]

[5] Fahse, T., Blohm, I., Hruby, R., and van Giffen, B. (2022). Explanation interfaces for sales forecasting. In *ECIS 2022 Research-in-Progress Papers*. [10]

[6] Fattah, J., Ezzine, L., Aman, Z., Moussami, H. E., and Lachhab, A. (2018). Forecasting of demand using arima model. *International Journal of Engineering Business Management*, 10:1847979018808673. [4]

[7] Giles Hooker, Hooker, G., Giles Hooker, Lucas Mentch, Mentch, L., Siyu Zhou, and Zhou, S. (2021). Unrestricted permutation forces extrapolation: variable importance requires at least one more model, or there is no free variable importance. *Statistics and Computing*, 31(6):1–16. [9]

[8] Gordon, A., Breiman, L., Jerome H. Friedman, Olshen, R. A., and Charles J. Stone (1984). classification and regression trees. *Biometrics*. [9]

[9] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Comput. Surv.*, 51(5). [8]

[10] Guolin Ke, Guolin Ke, Ke, G., Qi Meng, Meng, Q., Taifeng Wang, Wang, T., Chen, W., Wei Chen, Wei Chen, Wei Chen, Chen, W., Chen, W., Weidong Ma, Ma, W., Tie-Yan Liu, and Liu, T.-Y. (2017). A Highly Efficient Gradient Boosting Decision Tree. *Neural Information Processing Systems*, pages 3108–3116. MAG ID: 2753094203 S2ID: 497e4b08279d69513e4d2313a7fd9a55dfb73273. [5, 15]

[11] Hyndman, R. J. and Athanasopoulos, G. (2021). *Forecasting: principles and practice*. OTexts, 3 edition. [6]

[12] Ian Covert, Covert, I., Scott Lundberg, Lundberg, S., Su-In Lee, and Lee, S.-I. (2020). Understanding Global Feature Contributions With Additive Importance Measures. *Neural Information Processing Systems*, 33:17212–17223. [9]

[13] Ingle, C., Bakliwal, D., Jain, J., Singh, P., Kale, P., and Chhajed, V. (2021). Demand forecasting: Literature review on various methodologies. In *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–7. IEEE. [4]

[14] Jerome H. Friedman, Friedman, J. H., and Jerome H. Friedman (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232. MAG ID: 1678356000 S2ID: 1679bed-dda3a183714d380e944fe6bf586c083cd. [5]

[15] Kristin Blesch, David S. Watson, and Marvin N. Wright (2023). Conditional feature importance for mixed data. *AStA Advances in Statistical Analysis*. [9]

[16] Leo Breiman and Breiman, L. (2001). Random Forests. *Machine-mediated learning*, 45(1):5–32. [5]

[17] Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 4768–4777, Red Hook, NY, USA. Curran Associates Inc. [9]

[18] Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38(4):1346–1364. [5, 14, 17]

[19] Molnar, C. (2022). *Interpretable Machine Learning*. online, 2 edition. [8]

[20] Molnar, C., Gruber, S., and Kopper, P. (2020a). Limitations of interpretable machine learning methods. MAG ID: 3041627266. [9]

[21] Molnar, C., Molnar, C., König, G., König, G., Herbinger, J., Herbinger, J., Freiesleben, T., Freiesleben, T., Dandl, S., Dandl, S., Scholbeck, C. A., Scholbeck, C. A., Casalicchio, G., Casalicchio, G., Grosse-Wentrup, M., Grosse-Wentrup, M., Bischl, B., and Bischl, B. (2020b). Pitfalls to avoid when interpreting machine learning models. *arXiv.org*. [9]

[22] Rodrigo, J. A. and Ortiz, J. E. (2024). Global forecasting models: Dependent multi-series forecasting (multivariate forecasting). [6]

[23] Saluja, R., Malhi, A., Knapič, S., Främling, K., and Cavdar, C. (2021). Towards a rigorous evaluation of explainability for multivariate time series. pages –. [10]

[24] Scott Lundberg, Lundberg, S., Gabriel Erion, Erion, G., Hugh Chen, Chen, H., Alex J. DeGrave, DeGrave, A. J., Jordan M. Prutkin, Prutkin, J. M., Bala G. Nair, Nair, B. G., Ronit Katz, Katz, R., Jonathan Himmelfarb, Himmelfarb,

J., Flynn, J. T., Nisha Bansal, Bansal, N., Su-In Lee, and Lee, S.-I. (2020). From Local Explanations to Global Understanding with Explainable AI for Trees. *Nature Machine Intelligence*, 2(1):56–67. [9]

[25] Spiliotis, E., Makridakis, S., Semenoglou, A.-A., and Assimakopoulos, V. (2022). Comparison of statistical and machine learning methods for daily sku demand forecasting. *Operational Research*, 22(3):3037–3061. [4, 5]

[26] Tianqi Chen, Chen, T., Carlos Guestrin, and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *arXiv: Learning*. [5, 9]

[27] Vandeput, N. (2023). *Demand forecasting best practices*. Manning. [4, 6, 10]

[28] Vikas C. Raykar, Arindam Jati, Sumanta Mukherjee, Nupur Aggarwal, Kanthi K. Sarpatwar, Giridhar Ganapavarapu, and Roman Vaculín (2023). TsSHAP: Robust model agnostic feature-based explainability for time series forecasting. *arXiv.org*. ARXIV_ID: 2303.12316. [9]