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

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

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

Abstract

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

Supervisor: dr hab. Marek Sikora, Associate Professor

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 crucial for businesses, affecting inventory management, supply chain efficiency, and customer satisfaction. Demand forecasting issues are not limited 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.

In the thesis, the original taxonomy of demand forecasting issues was presented. Each task has been described with the specific problems that may be associated with each of them. This taxonomy can show a broader picture of tasks related to 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 topic of analysis and forecasting of promotion effectiveness was first considered. A set of indicators was proposed that can be used to evaluate promotions. The process of creating conditional attributes and a training set was shown to train predictive models for each indicator. Then it was presented how the proposed predictive models can be used in a promotional recommendation system and to discover knowledge about price sensitivity. Finally, a novel approach for analysing the quality of indicators using the original method for induction of Survival Action Rules was considered.

The next topic raised was the issue of splitting higher-level forecasts into lower-level forecasts. It is top-down forecasting. The comparison of selected forecasting algorithms was made. The possibility of using the *product types dissimilarity table* was presented, which allowed the product types (nominal values) to be changed to the distances used in the considered approaches. The conversion of attribute *size* into numerical characteristic was also proposed.

The next chapter presented the possibilities and advantages of using additional time series to improve the forecast performance. First, the possibility of using analogous time series from other locations was shown. Different possibilities for grouping and adding information from similar time series were presented. For both sets studied, it was shown that creating forecast models using additional data improved the results of the forecasts obtained. Next, a forecasting issue was examined, where additional time series from the same location were added to forecasting model. This approach in itself was innovative for the field under study, namely energy and electricity forecasting. The review of the literature indicated that this approach was unique. It

could be explored thanks to creating a digital-twin model of a building. By obtaining data on electricity consumption from household appliances and feeding this into the forecasting model, improvements in the forecasting of overall energy consumption were achieved. The conclusion of this research is that adding data from additional time series can significantly improve the quality of the forecasts.

The final case considered was the problem of forecasting product returns. Initially, a custom algorithm was proposed to combine the return transaction with the purchase transaction. The given algorithm was necessary to create a training set. As later results also showed, the addition of information about the purchase transaction and its characteristics significantly improved the prediction results. Good classification quality was obtained, confirming the effectiveness of the proposed process.

The presented thesis does not focus on the typical issue of time series forecasting in the sense that, given historical demand data, a forecast will be made for subsequent points in time. The focus was on lesser known demand forecasting issues. The research confirmed the thesis statement that the proposed taxonomy of demand forecasting problems allowed them to be resolved in a tailored way using domain knowledge and machine learning methods. Due to the cross-section of data used and issues addressed, it is hoped that the knowledge presented will be a valuable contribution to machine learning research and business communities. The research and dissertation were supported by the European Social Fund, grant InterPOWER (POWR.03.05.00-00-Z305).