

Improving Demand Prediction and Reducing Out-of-Stock – Application of Advanced Data Analytics in Retail Supply Chains

Brandtner Patrick¹, Darbanian Farzaneh¹, Falatouri Taha¹, Udokwu Chibuzor¹

1. University of Applied Sciences Upper Austria, Steyr, Austria

Corresponding author: patrick.brandtner@fh-steyr.at

Abstract: Correct demand prediction is a key success factor of efficient and demand-driven supply chains. This is especially true for the retail sector, where out-of-stock products directly influence customer satisfaction. In a data-driven world, advanced analytics approaches offer huge potential for demand prediction. The paper applies two of the most acknowledged demand prediction approaches to a real-world retail case. Based on an extensive database of cashier as well as Supply Chain data, we apply ARIMA and SARIMA and evaluate their applicability to predict the demand of selected perishable products. In addition, the impact of adding SARIMA-based demand forecasting to out-of-stock detection is analysed. The results show high applicability and a good forecasting quality especially of SARIMA. The quality of out-of-stock detection can significantly be improved by adding advanced analytics to traditional approaches in this area. For reaching higher demand prediction quality, results indicate the need to add the effect of promotions and the implications of substitute products to the applied approach.

Conference Topic(s): Area 2 - Distributed Intelligence in Physical Internet: ML and AI tools for demand forecasting and inventory control; Area 2 – Systems and Technologies for interconnected logistics: Machine Learning, Big Data and Artificial Intelligence.

Keywords: Data Analytics, Demand Prediction, Supply Chain Management, Big Data, Supply Chain Optimization, Value Network, Product Availability, Retail Management

1 Introduction

The retail sector is characterized by a high degree of demand uncertainty, volatile customer requirements and fierce competition by online retailers like Amazon. Having efficient and cost-optimized operational processes has become a key success factor for retailers, especially in the era of digitalization and the data-driven economy. Exact demand forecasting and optimal stock management are core abilities of every retail company and contribute significantly to efficient retail operations and subsequently to product availability and customer satisfaction (Pereira and Frazzon, 2020; Mandal 2020). Advanced data analytics based on Machine Learning (ML) and Artificial Intelligence (AI) offers huge potential to support and improve demand forecasting quality and correctness (Huber and Stuckenschmidt, 2020). However, for retail practitioners, the application of advanced data analytics is a challenging task and most approaches to demand prediction are done based on conventional statistical approaches and do not take into consideration additional factors on the level of the single stores and the specific conditions at this level (Lalou et al., 2020).

This research paper focusses on this issue and provides insights into the findings of a work package of a large research project conducted in the research field “Data Analytics and Foresight” at the department for logistics (Logistikum) at the University of Applied Sciences Upper Austria. The project partner is one of the main retailers in Austria with stores all across the country. The project’s aim is to implement data analytics in retail Supply Chains. More

precisely, the focused work package aims at improving the accuracy of demand prediction and out-of-stock detection for single stores by applying Machine Learning-based algorithms to optimize inventory control and demand forecasting for perishable items and products. The focal research question of the paper is hence defined as follows:

How can advanced analytics and Machine Learning-based approaches be applied in retail supply chains to improve demand forecasting and out-of-stock detection?

Based on an extensive analysis of cashier and Supply Chain data from the research partner company and for a selected region and a defined time span in Austria, the paper analyses two selected demand forecasting approaches and compares existing with improved out-of-stock detection methods. The remainder of the paper is structured as follows: section two provides an overview of forecasting methods and a discussion of innovative advanced analytics approaches and their benefits. Section three i) presents the data basis of the paper, ii) provides details regarding the approach used to fill missing data and iii) offers an overview of the selected products and their demand as the focus area of the paper. Section 4 discusses the results of applying two selected demand prediction approaches to the given retail case and shows the impact of this on out-of-stock detection. Finally, section 5 concludes the paper and provides an overview of its limitations and possible future research opportunities.

2 Demand Prediction and Out-of-Stock Detection in Retail SCM

Demand forecasting is one of the challenges that play an essential role in decision making in retail SCM (Da Fonseca Marques, 2020). However, it is not possible to calculate the real and actual customer demand in food retail stores because customers cannot place orders and can only buy products available at the point of sale. Therefore, product sales are considered as the real demand in the retail sector in general and in the current paper (Arunraj et al., 2016).

Inaccurate demand forecasting can cause understock or overstock problems. Understocking results in out-of-stock that ruins company's image and customers' trust. Overstocking does also bring negative effects on business, such as lack of storage space and waste of perishable products. Accurate demand forecasting helps businesses to increase their profit by managing their inventory and reducing out-of-stock. Data quality, data availability and forecast horizon are three factors that affect the accuracy of demand forecasting (Arunraj et al., 2016). Especially in the food retail sector, demand forecasting is crucial as perishable food products have short shelf lives and reducing both out-of-stock situations as well as outdated is important (Dellino et al., 2015).

Demand forecasting can be either short-term (between six to twelve months) or long-term (more than one year) and can be calculated with qualitative or quantitative methods. Quantitative methods are based on mathematical calculations. They aim to discover data patterns and provide accurate results. Machine learning techniques are among quantitative methods that can deal with big data with high processing speed and provide accuracy, adaptability and transparency (G and Prakash, 2020). Forecasting machine learning techniques can be classified into three groups. First, time series analysis such as autoregressive integrated moving average (ARIMA) and Holt Winters Exponential Smoothing (HW). Second, regression based methods such as Linear Regression (LR) and Support Vector Regression (SVR). And third, supervised and unsupervised methods like Support Vector Machine (SVM), Artificial Neural Network (ANN), Clustering and Hybrid modes (G and Prakash, 2020).

Time series analysis is the most popular approach for demand forecasting in retail industry due to its ability to capture trend and seasonality (Arunraj et al., 2016; Da Fonseca Marques, 2020). Shukla and Jharkharia (Shukla and Jharkharia, 2013) applied ARIMA models to predict daily

sales of onion in a wholesale market is India. Based on their results, Monday seems to be a day with higher demand than other weekdays for a certain product. Evers et al. (Evers et al., 2018) stated that ARIMA models are able to capture seasonal effect of demand and perform better than simple linear regressions. They compared tree models with linear regression models and proved that the first group outperforms the second with more than 99 percent accuracy to predict daily sales of bread in an online supermarket in Netherlands. Besides ARIMA, seasonal ARIMA (SARIMA) is one of the most used demand forecasting approaches in the retail context (Arunraj et al., 2016; Roque et al., 2019). The current paper focusses on time-series analysis and more precisely on applying ARIMA and SARIMA to the food retail context to improve demand prediction as well as out-of-stock detection.

3 Data Basis

In this section, the data basis, the approach used to clean and prepare the data for the analysis as well as an overview of the selected data is presented.

3.1 Data Basis of the Paper

The data available for the analysis covers more than 17 million records related to cashier data of three perishable products for more than 90 stores in Austria from January 2017 to December 2019. Sample data for this study consists of aggregated daily sales data of stores (cashier data) that were consistently open within this time period (66 stores). Therefore, the aggregated daily sales dataset for these three products consists of more than 170,000 records. In addition, around 10,000 calculated missed data elements (about 5% of total records) have been added to the database following the formula in section 3.2. In addition we used operational Supply Chain data to fill missing elements in cashier data when needed. This SC data mainly included shipping records from distribution center to stores.

3.2 Filling missed data

Running a reliable statistical analysis requires complete and full data. However, missing values are an inevitable situation in big data series. To fill missing data elements, we used alternative data from operational Supply Chain process (shipping data from distribution center to store) if available. If not available, there are several methods to complete missing data in time series analysis such as likelihood methods and neural networks or total random methods. In this research, we have used Simple Arithmetic Average (SAA Simple arithmetic average) (Duan et al., 2013).

$$\hat{Y}_{m_k t} = \frac{1}{N} (Y_{m_k 1} + Y_{m_k 2} + \dots + Y_{m_k N}) \quad (1)$$

In the formula (1), $\hat{Y}_{m_k t}$ indicates calculated missed data, N represent the number of available homologous days and $Y_{m_k N}$ are available data. If there is data missed in one special date in one year data, the homologous day for other years has been considered as the input of the formula to calculate estimation for the specific day.

In addition to missed data there are some noises in the data that could lead to forecasting results with lower accuracy. These noises may come either from external sources like promotions or holidays (Arunraj et al., 2016) or from logistic effects, bullwhip effects or temporal aggregation (Murray et al., 2018).

3.3 Overview of Data Selected for Demand Prediction

The data used in this paper focused on three selected perishable food items from the category of fruits and vegetables: i) cucumber, ii) salad and iii) avocado. The following figure provides an overview of the aggregated actual sales of the items in the selected 60 stores within the given time frame:

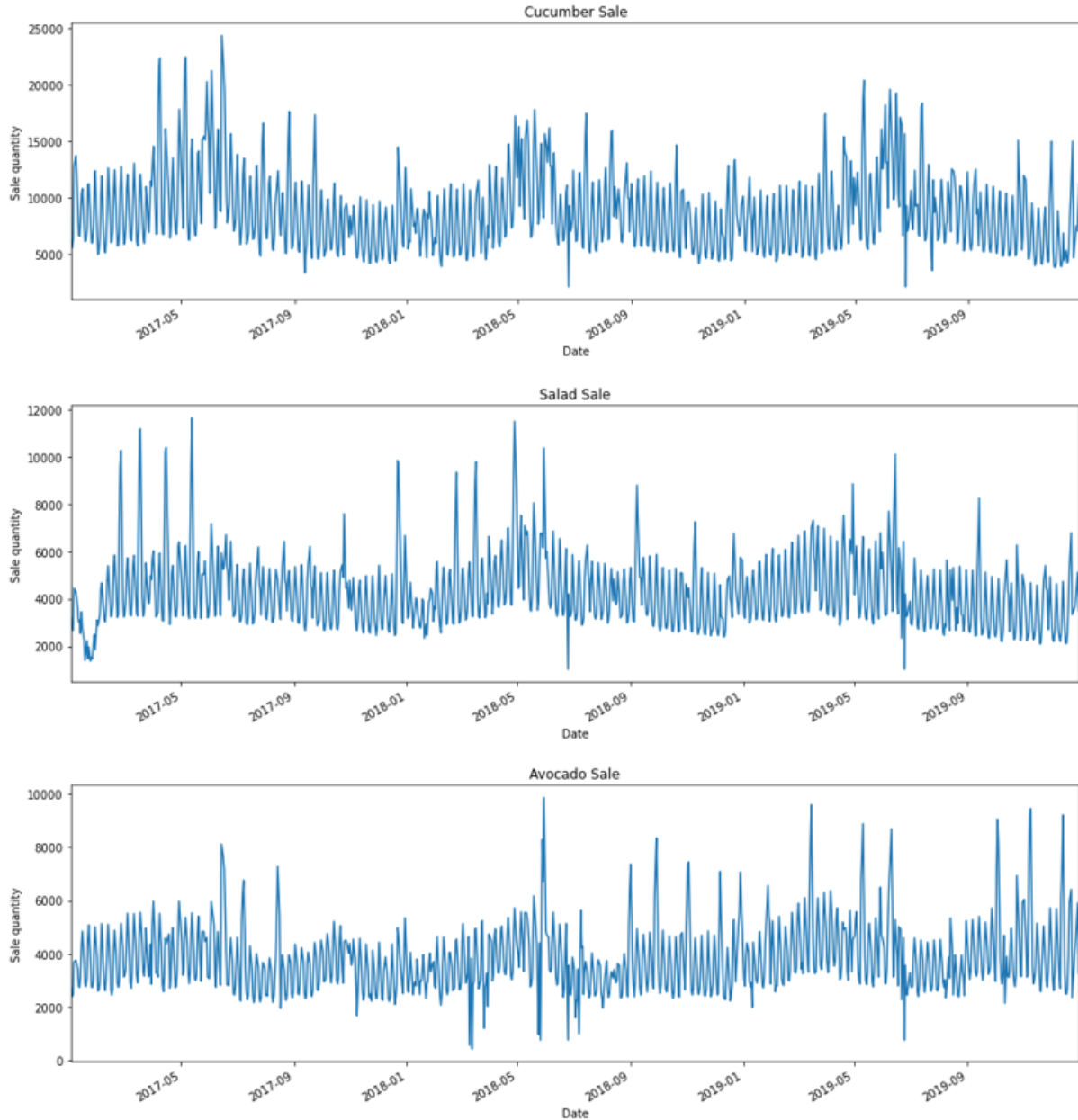


Figure 1: Aggregated actual sales of selected products

4 Results of Demand Prediction and Out-of-Stock Detection

Based on the selected products, ARIMA and SARIMA were analysed in terms of their applicability to predict demand. The results of the comparison of these two approaches is presented in section 4.1. Subsequently, section 4.2. discusses the effects of adapting out-of-stock detection based on the demand forecast by applying SARIMA.

4.1 Comparing ARIMA and SARIMA

To predict daily sales of selected stores in January 2020, the two most well-known time series algorithms ARIMA and SARIMA have been applied to historical sales data from January 2017 to December 2019. The two models have been evaluated with calculating the mean absolute percentage error (MAPE) (Lewis, 1982) based on the following formula:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|\hat{y}_t - y_t|}{y_t} \quad (2)$$

In the formula (2) n represents the total number of records, y_t shows the observed value at time t and \hat{y}_t indicates the predicted quantity. The difference between these two shows the deviation between reality and expectation. Dividing the summation of mentioned deviation shows the average deviation of expectation and reality. The less this number is the more accurate forecasting results are. The following table provides a framework for the interpretation of the calculated average (Lewis, 1982).

Table 1: Interpretation of MAPE.

MAPE	Interpretation
<10	Highly accurate forecasting
10-20	Good forecasting
20-50	Reasonable forecasting
>50	Inaccurate forecasting

In order to forecast the data we have used SPSS Modeler 18.2. with time series node. Table 2. illustrates the prediction results. The final model has been selected with Akaike (AIC) and Bayesian Information Criterion (BIC). As shown, SARIMA model performs better in all products with provided p , d and q parameters. The parameter p shows the number of time lag in auto regressive models (AR), d indicates the degree of differencing and q is the moving average order. For SARIMA, parameters P , D , Q represent the corresponding seasonal parameters to p , d and q . Although forecasting results are satisfying according to MAPE, there are some noises in the data which are not considered in our prediction model. This includes promotions and holiday effects that made irregular deviations from the fitted model. Hence, it is suggested to consider these parameters as external factors in future research.

Table 2: Comparing ARIMA and SARIMA results to predict sales from Jan. 2017 to Dec. 2019.

	AVOCADO	AVOCADO	CUCAMBER	CUCAMBER	SALAD	SALAD
	ARIMA	SARIMA	ARIMA	SARIMA	ARIMA	SARIMA
p, d, q	$p=0, d=0, q=7$	$p=3, d=0, q=3$	$p=0, d=0, q=7$	$p=0, d=0, q=4$	$p=0, d=0, q=10$	$p=0, d=0, q=2$
P, D, Q	-	$P=1, D=1, Q=1$	-	$P=1, D=1, Q=1$	-	$P=1, D=1, Q=1$
RMSE	1,023.306	909	2,506.811	2,094.040	1,102.381	953
RMSPE	50	48	34	24	32	27
MAE	730	544	1,865.378	1,346.918	807	608
MAPE	23	17	23	16	20	15
MAXAE	5,451.908	5,363.323	15,856.889	16,736.227	5,416.038	5,520.990

MAXAPE	678	762	443	241	427	372
AIC	13,009.143	12,704.178	14,688.995	14,258.716	13,148.781	12,790.952
BIC	13,043.049	12,733.202	14,718.058	14,282.902	13,182.688	12,810.302

As shown in table 2, SARIMA provides a better approach to forecast the demand of the three selected products. The MAPE for SARIMA was found to be 15 for salad, 16 for cucumber and 17 for avocado. Following Lewis (1982) and considering the values provided in table 1, the forecast quality of SARIMA can hence be considered as “good forecasting”. For ARIMA, MAPE is between 20 and 23, which represents a “reasonable forecasting” quality.

The following figure 2 compares the actual sales (blue line) with the predicted sales (orange line) based on the better performing SARIMA approach for cucumber, salad and avocados in January 2020. The high peaks for cucumber and salad (marked with red circles) represent the effect of promotions, which resulted in higher sales than forecasted:

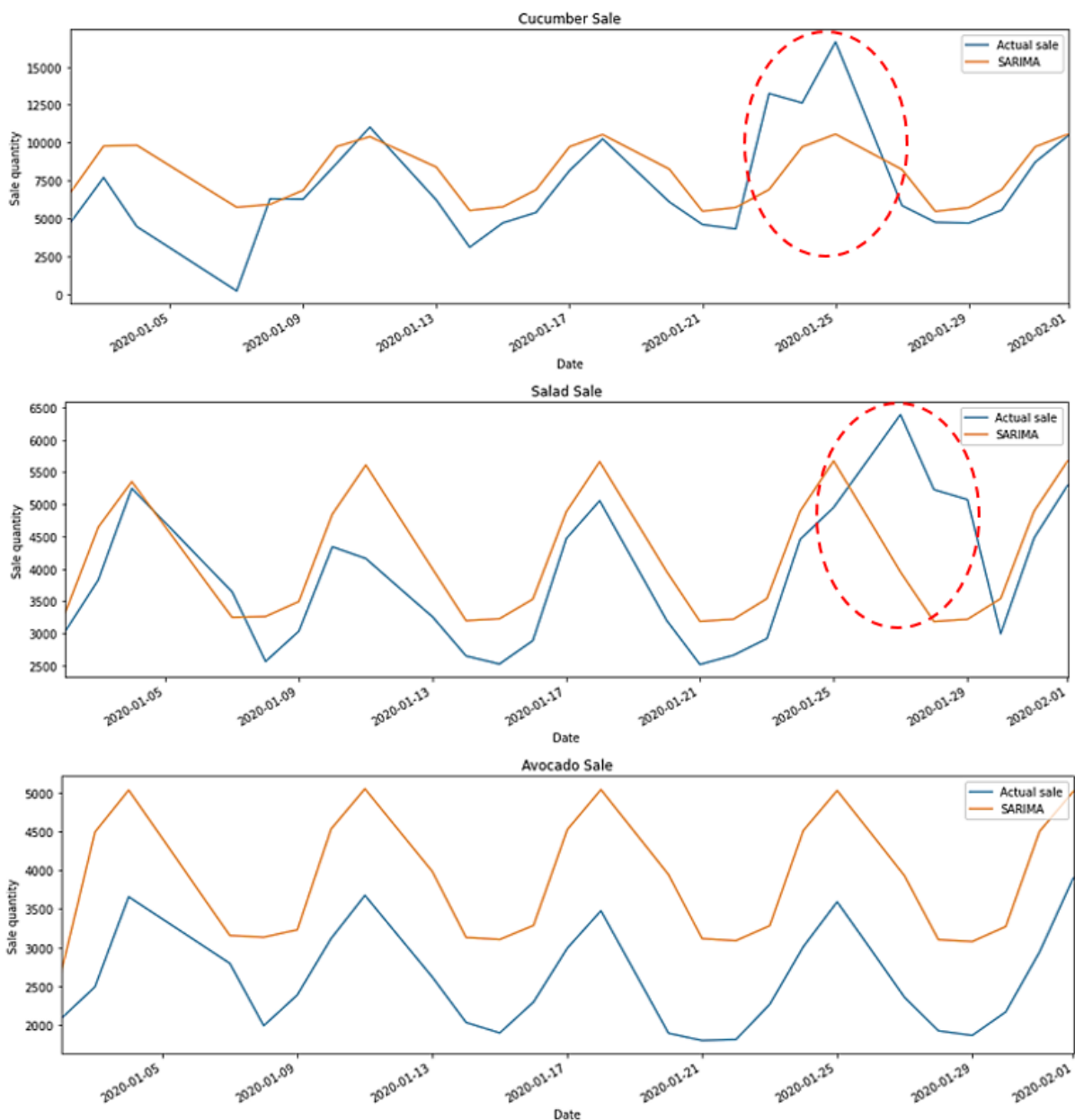


Figure 2: Predicted vs. actual sales of products in January 2020

Using the output of the forecasting model to test the data of the training records had satisfying results for cucumber and salad. However, for avocados (lowest visualization in figure 2) the results showed to be rather disappointing which raised the question of what external factors may have influenced the sale of this product in January. We paid particular attention to substitution effects between products as shown in table 3.

Table 3: Substitution effect of avocados.

Product	December 2019	January 2020
Avocado – sub1	20051	32569
Avocado – sub2	42082	46250
Avocado - target	103349	96141
Total	165482	174960

The table shows that while the total sale of avocados in January is higher than in the previous month, the trend of sale in the targeted product (i.e. the selected type of avocado) was declining. At the same time, alternative products (different pack-sized avocados as well as organic avocados) experienced a dramatic growth in sale. Hence, we suggest considering the replacement effect between products in future forecasting approaches.

4.2 Out-of-Stock Detection

Traditional methods for detecting out of stock situations were based on sale gaps in the last hours of a working day in a store. This approach has two main drawbacks: First, out-of-stock situations are considered to be the result of insufficient supply from distribution center to store only. Operational out-of-stock factors at the store level may not be detected, e.g. the fact that store personnel didn't manage to put products from storage to shelf. This is especially the case for product unavailability at shelves during the working hours. Second, the traditional out-of-stock detection method only works as a flag which indicates if an out-of-stock occurred or not. It does not allow for estimating how many product sales were potentially missed.

Based on the forecasted sales minus MAPE we identified out of stock situations as well as their amount. More precisely, the differences between expected sales and real sales (if higher than the average amount of MAPE) were used to determine out-of-stock and their respective amount. As shown in table 4, the proposed model identified more out-of-stock situations than the traditional approach. However, the main advantage of this adapted method is that it allows for estimating missed sale quantities.

Table 4. Out of Stock results for one selected store for January 2020.

Product	Out-Of-Stock situations Traditional method	Out-Of-Stock situations Forecasting method	Missed sale	Ave daily sale
Cucumber	1	2	221	157
Salad	1	9	168	79
Avocado	1	5	104	45

5 Conclusions and Outlook

The results of the paper indicate that there are big benefits of applying advanced analytics in retail demand forecasting and out-of-stock detection. Compared with the traditional out-of-

stock detection approaches used in the given retail case, the applied improved model identified a far higher amount of critical out-of-stock situations at the point-of-sale. The improved out-of-stock detection approach does not only focus on the overall sold quantity and the sales gaps (i.e. times without sales) at the last hours of the day but also includes the forecasted daily sales of stores. Regarding the improvement of demand forecasting, time series forecasting and more precisely auto-regressive integrated moving averages (ARIMA) and seasonal auto-regressive integrated moving averages (SARIMA) was used to better predict the sales of the single stores on daily levels. In general, the results show good applicability of the applied approaches, especially SARIMA seems to be applicable and offers a good forecasting quality.

The practical implications of the paper are as follows: first, practitioners get an overview of how demand forecasting and out-of-stock detection is currently conducted in retail SCM practice. Second, innovative advanced analytics and Machine Learning-based approaches for improving these two tasks are discussed, providing practitioners with an overview of potential future improvements. Third, an actual practice use case from the Austrian retail sector is provided as a best practice example on how to improve demand forecasting and out-of-stock detection by means of Machine Learning-based approaches. This best practice example shows how advanced analytics can be implemented and adopted to a real-life use case in the retail sector. From a scientific point-of-view, the paper provides new insights in terms of presenting and discussing the applicability of time series forecasting with ARIMA and SARIMA in retail practice. For distributed intelligence in physical internet setting, the results of the paper provide the basis for establishing a more efficient demand planning across the network of retail stores, retail distribution centers, regional as well as trans-regional product suppliers and logistics serviced providers. By sharing the forecasted demand levels in an interconnected logistics network, out-of-stock situations, food wastage and unfulfilled customer demands can be reduced respectively avoided. In addition, transportation as well as warehouse infrastructure can be planned closer to real demand at an early stage, resulting in a better utilization of resources and the avoidance of express orders and shipping. Ultimately, these effects also contribute to lower emissions and lower product wastage due to expired items.

The limitation of the paper is the focus on one specific retail case in Austria. However, by having an extensive data base from actual sales at the point-of-sale as well as data along the retail Supply Chain provided a solid basis for data analysis and demand forecasting as well as out-of-stock detection improvement and evaluation. Future research should focus on adding and analysing the effect of substitute products as well as the effect of promotions on SARIMA-based prediction approaches.

Acknowledgements

This research was partly funded by the government of Upper Austria as part of the Excellence Network Logistics – Logistikum.RETAIL.

References

- Arunraj N. S., Ahrens D., Fernandes M. (2016): *Application of SARIMAX model to forecast daily sales in food retail industry*. International Journal of Operations Research and Information Systems (IJORIS), v7, no2, 1–21.
- Chen K.; Wang C. (2007): *A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan*. Expert systems with applications, v32, no1, 254–264.

- Da Fonseca M., Rúben A. (2020): *A Comparison on Statistical Methods and Long Short Term Memory Network Forecasting the Demand of Fresh Fish Products*. PhD-Thesis, Universidade do Porto.
- Dellino G., Laudadio T., Mari R., Mastronardi N., Meloni C. (2015): *Sales Forecasting Models in the Fresh Food Supply Chain*. Proceedings of ICORES 2015, 419–426.
- Duan W., Cao Q., Yu Y., Levy S. (2013): *Mining Online User-Generated Content: Using Sentiment Analysis Technique to Study Hotel Service Quality*. Proceedings of the 46th Hawaii International Conference on System Sciences: IEEE, 3119–3128.
- Evers J. M., Tavasszy L., van Duin R., Schott D., Gorte F. (2018): *Demand forecast models for online supermarkets*. Proceedings of Nectar Cluster 2: E-groceries, digitalization and sustainability: Which governance, planning and regulation mix do our cities need?, 12.
- G S.F., Prakash N. (2020): *Machine Learning in Demand Forecasting - A Review*. Proceedings of the 2nd International Conference on IoT, Social, Mobile, Analytics & Cloud in Computational Vision & Bio-Engineering (ISMAC-CVB 2020), 9.
- Huber J.; Stuckenschmidt H. (2020): *Daily retail demand forecasting using machine learning with emphasis on calendric special days*. International Journal of Forecasting, v36, no4, 1420-1438.
- Lalou P., Ponis S.T., Efthymiou O. K. (2020): *Demand Forecasting of Retail Sales Using Data Analytics and Statistical Programming*. Management & Marketing, v15, no2.
- Lewis C. D. (1982): *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. Butterworth-Heinemann.
- Mandal P. C. (2020): *Retailing Trends and Developments - Challenges and Opportunities*. International Journal of Business Strategy and Automation (IJBSA), v1, no2, 1–11.
- Murray P. W., Agard B., Barajas M. A. (2018): *ASACT-Data preparation for forecasting: A method to substitute transaction data for unavailable product consumption data*. International Journal of Production Economics, v203, 264–275.
- Pereira M. M., Frazzon E. M. (2020): *A data-driven approach to adaptive synchronization of demand and supply in omni-channel retail supply chains*. International Journal of Information Management, 102165.
- Roque L., Fernandes C. A. C., Silva T. (2019): *Optimal Combination Forecasts on Retail Multi-Dimensional Sales Data*. arXiv preprint arXiv:1903.09478.
- Ruiz-Aguilar J. J., Urda D., Moscoso-López J. A., González-Enrique, J., Turias I. J. (2020): *Container Demand Forecasting at Border Posts of Ports: A Hybrid SARIMA-SOM-SVR Approach*. In : Bernabé Dorronsoro, Patricia Ruiz, de la Torre, Juan Carlos, Daniel Urda, El-Ghazali Talbi (Eds.): Optimization and Learning. Cham: Springer International Publishing, 69–81.
- Shukla M., Jharkharia S. (2013): *Applicability of ARIMA models in wholesale vegetable market: An investigation*. International Journal of Information Systems and Supply Chain Management (IJSSCM), v6, no3, 105–119.
- Svetunkov I., Boylan J. E. (2020): *State-space ARIMA for supply-chain forecasting*. International journal of production research, v58, no3, 818–827.