Optimizing Demand Forecasting: Classical Statistical Models vs. AI-Driven Approaches

Coralia ZOTIC*
Bucharest University of Economic Studies, Bucharest, Romania
**Corresponding author, tanasuicacoralia22@stud.ase.ro*

George DINU*
Bucharest University of Economic Studies, Bucharest, Romania
**Corresponding author, george@stud.ase.ro*

Mihai-Daniel ROMAN
Bucharest University of Economic Studies, Bucharest, Romania
mihai.roman@ase.ro

Abstract. *International competition between retailers is very intense in actual world and the capacity to remain profitable is essential influenced by the ability to forecast the demand as accurately as possible. Also, is very important to minimize stock outages, to reduce the volume of each product in stock, to reduce costs and to increase the revenues. The current work proposes a new approach to forecasting and optimization in the retail environment and introduces an improved method for identifying the optimal algorithm that can help a retailer to get as close as possible to mentioned conditions. The research methodology identifies the optimal forecasting model and then optimize the model results and parameters such as the optimal order quantity, the reordering point, the maximum stock and the safety stock. To obtain the optimal solutions we develop distinct forecasting and optimization algorithms, starting with classical econometric methods like ARIMA, or dynamic optimization, and progressing to more complex approaches, including transformers, neural networks, Fourier transformation, or combined algorithms like Fourier transformation with Prophet. An important challenge was identifying an optimal method for selecting the best operating algorithm, as classical methods like RMSE, MAPE, or MAE failed to identify the most accurate prediction algorithm. In order to test the algorithm's quality, we test twenty-six distinct forecast algorithms and compared at each product level, identifying Fourier transformation, either on its own or in combination with other algorithms, as a highly effective forecasting model. Finally, the Continuous Ranked Probability Score is proposed as the best model selection strategy.*

Keywords: Machine learning, Artificial Intelligence, Genetic Algorithm, CRPS, Inventory management, forecasting, optimization

Introduction

Competition between retailers is intense, and to be able to remain in the game while remaining profitable, it is essential to forecast the demand as accurately as possible. This ensures, on one hand, the ability of being able to meet the demand and on the other hand to minimize the stocking costs as much as possible. All these strategies aim to maximize the efficiency of the invested capital and minimize capital costs through effective stock optimization.

The current work proposes a different perspective on demand forecasting arguing that relying on a single algorithm, even if it is a complex architecture of algorithms, is unlikely to achieve higher accuracy compared to training multiple algorithms at the product level and selecting

DOI: 10.2478/picbe-2025-0103

the most suitable one for each product. By following this approach, the final forecast for the entire product population will be the result of distinct models running in parallel for each product.

The second stage of the research proposes the following workflow: the forecasted demand at the product level will serve as one of the input variables of an optimization algorithm. This algorithm aims to determine the essential parameters for efficient inventory management, including the optimal safety stock, reorder point, optimal order quantity, and maximum stock. By using this optimization algorithm, the retailer can avoid stock outages and identify the optimal maximum stock to effectively address overstock issues.

The case study was based on a database consisting of the monthly sales data at the product level for a sample of 15.000 distinct products. These products were selected from the categories of sanitary installations, heating systems, and other related installation products. The historical data covered a time horizon of 2 years and 8 months, starting in January 2022 and ending in October 2024. The forecasting was conducted for 12 future data points, corresponding to 12 months. Initially, an attempt was made to identify the best forecasting model at the product level using standard metrics such as RMSE, MAPE, and MAE. However, after a detailed visual evaluation of the forecasted curves generated by different models at the product level, some models were identified as the best-performing or most accurate. Yet, for example, these models forecasted only a constant line, despite the historical data showing dynamic patterns with high amplitude and significant variation from month to month. Other distinct types of scores were developed to identify the best model, and numerous tests were conducted. Still, finally, the Continuous Ranked Probability Score (CRPS) was recognized as the best selection strategy for determining the most accurate predictive model. A total of 24 distinct algorithms were trained for forecasting the demand, and 10 distinct algorithms were used for parameter optimization. The best algorithm for optimizing the parameters was a genetic algorithm.

The research also includes an exploratory data analysis aimed at identifying whether certain algorithms perform better on specific types of products and determining the characteristics of those products. This is an important step, as running numerous distinct models can result in inefficient use of computational resources. Therefore, it is necessary to optimize the selection and application of forecasting models, taking into account the computational resources required by each algorithm.

The research is structured into five chapters, with the primary aim of improving the overall performance of forecasting models and better detecting the optimal parameters for inventory management. The second chapter provides a summary of relevant literature, focusing on machine learning models used for previsioning sales or demand, as well as strategies for evaluating the models and optimizing inventory parameters. The next sections describe the methodologies and the models applied in the case study, present the conclusions of the research, outline its limitations, and propose ideas for further development. The final section includes the bibliographic resources used.

This research proposes a solution to a key concern for retailers: how to forecast demand more accurately in order to avoid stockouts and better optimize inventory levels, ultimately reducing the company's debt exposure. The key objectives of the study are to (1) test a wide range of forecasting models and identify the best-performing one for each product individually, (2) improve the model selection methodology by replacing classical accuracy metrics like RMSE or MAPE with CRPS, and (3) optimize inventory parameters such as reorder point, order quantity, and safety stock using a genetic algorithm. The uniqueness lies in its multi-model approach, testing 26 forecasting algorithms per product and selecting the best one based on CRPS, as well as in integrating forecasting with a dynamic inventory optimization algorithm to ensure both accuracy and cost-efficiency in retail operations.

Literature review

Forecasting Methods

There are plenty of studies analyzing and comparing different forecasting methods and their positive impact on a company's turnover, conditioned by achieving the highest possible accuracy. The primary objective for many researchers is to identify the model that performs best for the analyzed set of products or product segments, minimizing estimation errors.

In parallel, some researchers are focused on identifying optimal methods and methodologies to optimize inventory, with the primary goal of minimizing stock gaps or overstock issues. A third group of researchers aims to combine these two areas of study into a single research direction, recognizing that effective stock optimization is closely linked to a robust demand forecasting strategy at the product level.

Since the majority of studies focus on one of these two directions (forecasting or optimization), research on the proper way or method for selecting the best forecasting model is limited. Moreover, most studies rely on classical methods for selecting the optimal model, such as Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), or Mean Absolute Error (MAE). Fewer studies explore and utilize more advanced methods, such as Continuous Ranked Probability Score (CRPS), for identifying the optimal model.

An interesting methodology used to forecast demand is hierarchical time series forecasting (Abolghasemi, 2024). This methodology matches perfectly with some industries, especially companies operating in the fast-moving consumer goods (FMCG) sector, and less in others. Considering the retailer analyzed for my case study, this approach won't be sustainable because the retailer does not sell FMCG but its products through a B2B2C (business-to-business-to-consumer) system. This means promotions and campaigns have a low impact on sales.

Other research with a high degree of impact in forecasting methodologies, a direction that I also tested in my case study, proposes the usage of an assembly of models, combining the Fourier transformation with the so-called grey models. These grey models are defined as a class of mathematical models that perform in grey conditions, considering that there is a lack of complete information (black) and perfect information (white). They also perform well with limited and incomplete information (Ye, 2024). The results prove superior performance, subclassing complex models like neural networks (such as LSTM) or classical statistical models like AutoRegressive Integrated Moving Average (ARIMA) or Holt-Winters. An essential element contributing to the good performance of this assembly of models is directly connected to the structure of the input data. It performs well on non-stationary datasets, with low dimensions but high complexity from the seasonality perspective.

Forecasting time series is a complex task that does not provide an ideal or unique model for addressing it. Some researchers identify decision tree models as performing best (Long, 2025), achieving the lowest prediction errors (Wellens, 2024). Other studies demonstrate that neural networks, such as LSTM (Wu, 2024) or Gated Recurrent Unit (GRU) (Sukolkit, 2024), have the highest performance rates because they excel in capturing the complex dynamics of temporal series. However, some researchers prove that the optimal model is an ensemble of models, combining decision trees with neural networks (Ahmed, 2024), or other advanced analytics techniques that integrate machine learning with deep learning, such as the Neural Hierarchical Interpolation for Time Series (NHITS) model (Chae, 2024). These diverse results demonstrate the necessity of adapting methodologies to the specificity of the analyzed data.

Optimal model selection strategies

The classical methods for evaluating a forecasting model are RMSE, MAPE, and MAE. These measures effectively capture the deviation of the forecasted data from the actual data in the time series. However, if the goal is to identify the model that best aligns the prediction with the real values in terms of the dynamics of evolution—meaning that the variation in the forecast closely matches the historical variation of the time series—then the best methodology becomes CRPS (Berrisch, 2023).

Some studies identify the limitations of classical methods and propose CRPS as an alternative for identifying extreme events (Taillardat, 2023). Another study highlights the limitations of using the CRPS methodology, as the derivative method, CRPS-Sum, can produce **misleading** results **when** applied to datasets of higher dimensions (Koochali, 2022).

Optimization models and strategies

Even if the demand forecasting part is an essential one, it is not sufficient to achieve the objective of cost reduction and turnover increase through better demand identification and more accurate stock optimization. There is also a need to detect the optimal cybernetic models that will help companies reach this objective. Relevant research in this direction (Corsini, 2024) proves the efficiency of using an ensemble of cybernetic methodologies. This ensemble combines a simulation model, used for forecasting the future performance of industrial systems, with artificial neural networks (ANN) employed as predictive models. These are used to predict two essential factors for the performance of production and distribution systems: the fill rate and the average inventory level. The third element of the designed cybernetic system presented in the research is an optimization algorithm, particle swarm optimization (PSO), used to optimize the resupply parameters.

An interesting article (Younespour, 2024) focuses on presenting the increase in performance when optimizing inventory parameters through an ensemble of two optimization algorithms: PSO and Genetic Algorithm (GA). The study successfully demonstrates a significant reduction in the convergence time of the hybrid algorithm compared to the individual usage of the algorithms. It also proves the increase in the quality of solutions obtained by the combined algorithm compared to the solutions obtained when the models are used separately.

Another research in the optimization field, compares several optimization algorithms for inventory parameter optimization, aiming to identify the most profitable model, not only in terms of the precision of the results but also from the perspective of resource efficiency when running the algorithms (Keswani, 2024). The study identifies the Cuckoo Search Algorithm (CSA) as the most efficient one because it demonstrated the best performance, quick convergence, and stable results.

Neural networks remain the most widely used and tested algorithms for optimizing inventory parameters. An innovative approach adopted by a group of researchers (Dalal, 2024) demonstrates that combining two neural network algorithms to optimize the inventory process can increase sustainability, reduce costs, and enhance turnover. Their method integrates both temporal and spatial perspectives: a Convolutional Neural Network (CNN) for optimizing delivery routes and a Bidirectional Long Short-Term Memory algorithm (BiLSTM) for forecasting future demand, reducing stock outages, and avoiding overstock.

Another distinct perspective in forecasting optimization involves using algorithms to fine-tune the parameters of predictive models. For instance, PSO (Particle Swarm Optimization) is employed by Wang (2024) to optimize the parameters of a Support Vector Regression (SVR) model, with the ultimate goal of maximizing forecast accuracy for energy demand. The proposed architecture incorporates an Ensemble Empirical Mode Decomposition (EEMD) structure to address data non-stationarity. Following this, the Sample Entropy method simplifies the dataset into a more manageable but relevant time series, preserving essential signals for forecasting. SVR is then applied for demand forecasting, with its parameters optimized via PSO.

As already noted, there is no perfect forecasting methodology to suit all possible cases. However, an alternative involves using a commuting system that trains multiple algorithms and identifies the one that best serves each time series (Ahmed, 2024). This method is known as the Switching-Based Forecasting Approach (SBFA).

Some authors propose a holistic approach to address all the industrial aspects of a production company: demand forecasting, sales estimation, route optimization, inventory management, risk balancing, quality control, and product inspection, by using computer vision to analyze images. In this respect, they employ Convolutional Neural Networks (CNN) for quality checks and product inspection (Khedr, 2024).

The general improvement in the accuracy of the forecasted results recommends the use of these combined algorithms for improving the general performance of the economic activity.

Methodology

There are three important stages in the research that will be distinctly described in the following rows. Different strategies were applied, having the final objective of identifying the best direction for optimizing the inventories and reducing costs. For a better perspective on the flow of the study, a diagram was built (Figure 1), showing the succession of the actions taken throughout the research.

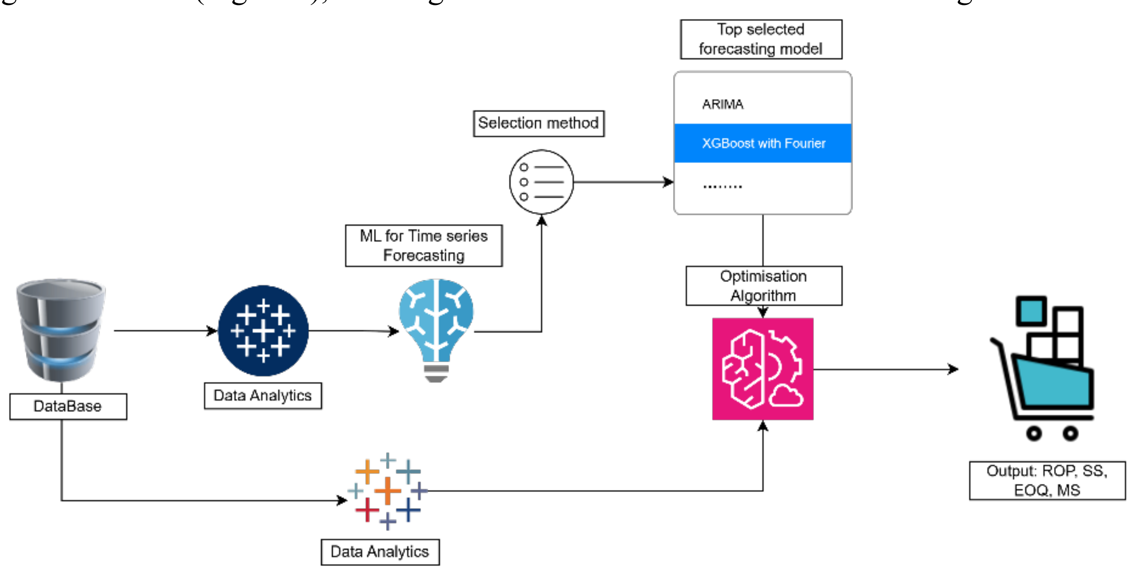

Figure 1. The actions flow and methodologies

Source: Author's contribution, generated using https://app.diagrams.net/.

The first part of the research is allocated to identifying the best forecasting algorithm. There were many time series algorithms trained to detect the best matching algorithm for each specific time series dataset. Even if not all algorithms were initially created for time series forecasting, they were adapted to this specific data. Five distinct classes of algorithms were tested: statistical models, classical machine learning models, advanced AI and deep learning models, hybrid and decomposition-based models, and seasonality and frequency-based techniques.

The group of statistical models includes: Autoregressive Integrated Moving Average (ARIMA), Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), Seasonal Autoregressive Integrated Moving Average (SARIMA), and Fourier-based Forecast. AR predicts future values using a linear combination of past values, while MA models the error term as a linear combination of past error terms. ARMA combines AR and MA to handle both past values and errors and ARIMA includes differencing into the ARMA model to handle non-stationary data. SARIMA adds seasonal components to ARIMA to capture seasonal patterns. Fourier-based Forecast is a statistical model that decomposes a time series into its frequency components using the Fourier Transform. It identifies the most significant seasonal patterns by retaining only the top frequencies and removing the rest. The forecast is generated by reconstructing the series based on these filtered frequencies and extrapolating future values. This approach efficiently captures and forecasts periodic or seasonal patterns in the data.

The class of classical machine learning models includes the following tested algorithms: linear regression with or without Fourier, LightGBM with Fourier, CatBoost with or without Fourier and XGBoost with or without Fourier. The last three algorithms are boosting algorithms, which enhance the performance of classical decision tree models by combining multiple weak learners into a strong predictive model through iterative training and error reduction. The time series data were adapted by generating lagged features to capture temporal dependencies, adding a month variable to represent seasonality, and incorporating Fourier terms (where applicable) to model periodic patterns, ensuring compatibility with the input requirements of these machine learning algorithms. The key features for all these models are that for the models with Fourier elements, they capture periodicity using sine and cosine terms and for the models without Fourier elements, they only rely on the lag features and simple seasonality indicators like the month variable.

The class of advanced AI and deep learning models were tested to identify their performance on distinct sets of time series. This class of algorithms contains: Long Short-Term Memory (LSTM), Transformer-Patch (TP), TSMixer (TSM), N-BEATS (NB), N-HITS (NH), and Kolmogorov-Arnold Inspired Neural Network (KAINN). These models preprocess the input data by normalizing the target variable, generating a sliding window to capture temporal dependencies, and feeding these sequences into a neural network to extract patterns for iterative forecasting. LSTM is designed for sequential data and captures both short- and long-term temporal patterns. TP algorithm combines multi-head attention for global dependencies, convolutional layers for local feature extraction, and dense layers for prediction. TSM uses one-dimensional convolutional layers for short-term dependency extraction and global average pooling for feature aggregation. NB algorithm is a fully connected deep neural network that directly predicts future values without explicit decomposition of components. NH combines convolutional layers for feature extraction and dense layers for interpolation and prediction, while KAINN is a deep fully connected neural network inspired by the Kolmogorov-Arnold representation theorem, with layers capturing complex non-linear dependencies. The Kolmogorov-Arnold representation theorem states that any multivariate continuous function can be decomposed into a finite sum of univariate continuous functions, enabling the modeling of complex, high-dimensional relationships using simpler components.

The Hybrid and Decomposition-Based class of models combines statistical, machine learning, and decomposition techniques to improve forecasting. The tested algorithms are: Prophet, Prophet with CatBoost (PC), Prophet with Linear Regression (PLR), Prophet with GARCH (PG), ARIMA with GARCH (AG), and Fourier with Spectral Analysis (FSA). Here's an overview of the models: Prophet decomposes time series into trend, seasonality, and holiday components; PC combines Prophet's trend and seasonality modeling with CatBoost which captures non-linear dependencies in residuals having the objective to correct residuals. PLR uses Prophet for trend and seasonality, and Linear Regression corrects residuals using a simple linear approach. PG combines Prophet's forecasts with GARCH, which models volatility in residuals. AG model combines ARIMA, which captures trends and seasonality, with GARCH which models variance fluctuations. The FSP algorithm uses the Fourier Transform to decompose a time series into its frequency components, retaining significant patterns for reconstruction and forecasting, while the spectral analysis while is the step of filtering and selecting the dominant frequencies to refine the forecast.

The second part of the study is consistently smaller in scope but equally important as the other parts and focuses on finding the best methodology to select the top-performing algorithm for a specific dataset. The classical approach involves using RMSE, MAPE, and/or MAE. RMSE is sensitive to outliers with large values, MAE is affected by very small values, and all three measures focus on point predictions rather than the dynamics of the prediction. CRPS was ultimately adopted as the final methodology, as it proved helpful in identifying the model that best aligns the historical evolution with the forecasted one.

Although CRPS was selected as the main model evaluation metric due to its ability to capture the dynamics of the forecast distribution, other selection strategies exist. Bayesian model averaging and probabilistic ensemble approaches are valuable alternatives, particularly when models are combined rather than selected individually. However, these methods often require higher computational resources and more complex calibration procedures, making them less practical when evaluating large numbers of models across thousands of product-level time series. In this context, CRPS provides a more scalable and interpretable option for individual model selection.

In the final part of the research, a genetic algorithm, combined with dynamic calculations was tested. The scope is to optimize the key inventory parameters: safety stock, reorder point (ROP), optimal reorder quantity, and maximum stock by reducing stock outages while also minimizing costs. The parameters are dynamically adjusted based on historical and forecasted demand, lead time, and cost factors. The approach includes a fitness function, that evaluates the total cost, and the genetic algorithm that iteratively improves the inventory policies through selection, crossover, and mutation.

Results and discussions

The database contains private data, consisting of 20.000 distinct Stock Keeping Units (SDKs) from the retail domain. It comprises monthly time series for each of these products, starting from January 2022 until October 2024. The products are already clustered by the company into distinct categories, based on product margin combined with product sales frequency. For better visibility, the product clusters will be denoted as follows: class H, products with high frequency and high margin (34 data points); class M, products with medium sales frequency and medium margin (more than 24 but fewer than 34 data points); class L, products with low frequency and low margin (fewer than 24 data points).

The dataset consists of monthly sales data provided by a large retailer active in the technical and construction goods market. Each product has a different volume of available data points, and this impacts the performance of the models trained on that specific amount of data. Complex neural network models and hybrid approaches tend to achieve better accuracy when more historical data is available, while statistical models score better when data points are limited. Promotional activity or marketing campaigns were not included in the dataset, as these factors have minimal influence within the business model analyzed.

In Figure 2 are presented the distributions of the top model per product and category. For the H class of products, 18 models surpass the baseline model in terms of the number of products for which they are chosen as the best. One of these 18 models was selected as the best algorithm for 91% of the products in cluster H. For the M segment of products, only 10 models were selected as the best for a broader range of products than the baseline model, covering just 63% of the products in the cluster. For the L segment of products, 11 algorithms were identified that outperform the baseline algorithm. 68% of the products in this segment selected one of these 11 algorithms as the most accurate in predicting future sales.

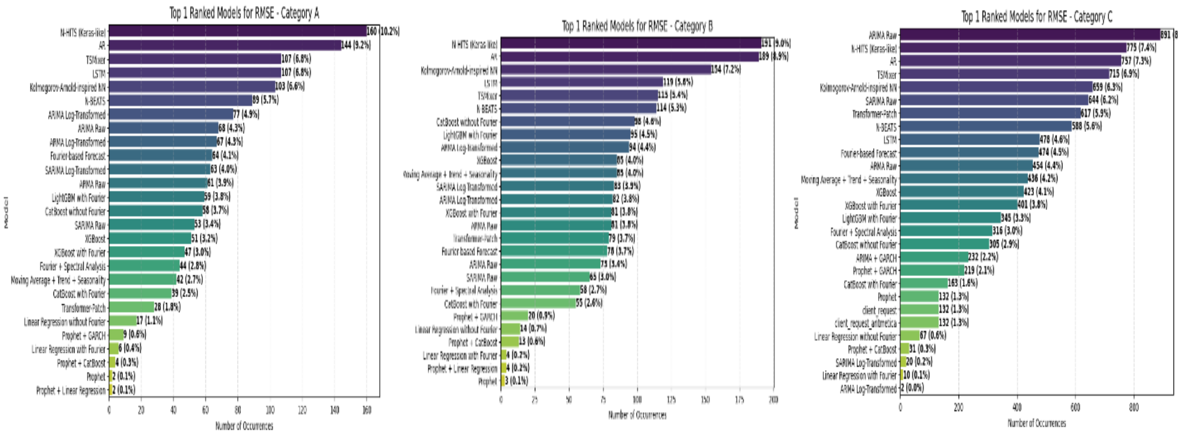

Figure 2. Best algorithm number of occurrence per class of products

Source: Author's contribution, generated using Python.

An important step in detecting the efficiency of using multiple forecasting algorithms is the analysis, at each product level, of the model with the lowest identified error score and its comparison with the baseline model. The goal is to identify, at the segment level, how many products have an optimal model that is different from the baseline model. Figure 3 is suggestive and provides the following insights: for the H class, only 2.7% of the products perform better with the baseline model than with any of the other models. For the M class, in only 4% of the products, no other model performs better than the baseline model. For the L class, 22.3% of the products have the baseline model as the top model in terms of the lowest error score.

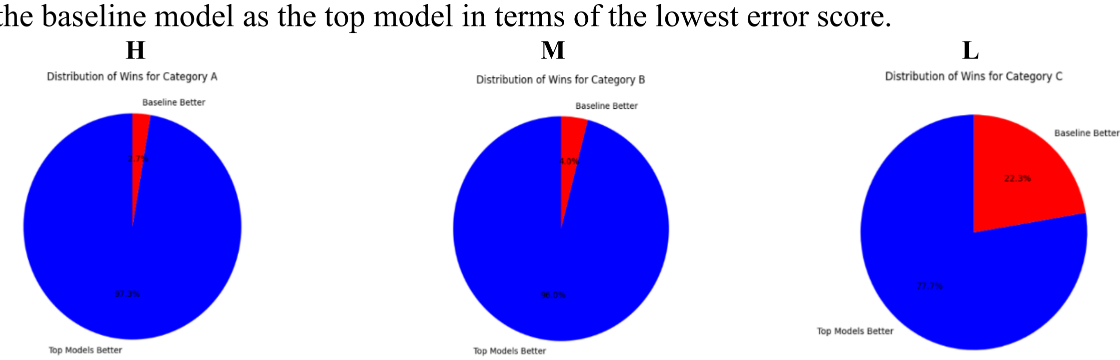

Figure 3. Products with best model other than the baseline model

Source: Author's contribution, generated using Python.

The models perform different along distinct product categories. That can be explained by the characteristics of the time series in each group. For high-frequency and high-margin products (class H), the availability of more historical data points and more stable sales patterns allows complex models—especially Fourier-based and hybrid approaches—to capture seasonality and trends more effectively. In the medium group (class M), where time series are slightly shorter and sales patterns less stable, models such as boosting algorithms (e.g., LightGBM with Fourier) perform well by balancing flexibility and generalization. For low-frequency products (class L), where the data is limited and more irregular, simpler statistical models like ARIMA or MA tend to outperform AI-based models, as they are better suited to small-sample forecasting and do not overfit. These differences highlight the importance of adapting forecasting techniques to the data profile of each product segment, rather than applying a single model uniformly.

But which are the predictors that influence or determine whether a model is more suitable for a time series than the baseline? To find the first insights in answering this question, a supervised machine learning model was built, with a binary target variable: 1 if the best algorithm has a lower RMSE than the baseline model and 0 if the chosen algorithm is the baseline itself. The predictors include Stationarity, Trend, Seasonality at 12 months, Seasonality at 6 months, Seasonality at 4 months, Seasonality at 3 months, Seasonality at 2 months, Adjusted CRPS Score, the number of historical data points, RMSE, and the Model name. Except for Adjusted CRPS Score, RMSE, and the historical data points, all other predictors are categorical variables. This determined the choice of the CatBoost algorithm, which is particularly effective in handling categorical variables. The Adjusted CRPS Score differs from the original CRPS formula due to a different penalty structure on the evolution of the forecasted series: if the forecast overestimates, the penalty is lower compared to when the forecast underestimates sales.

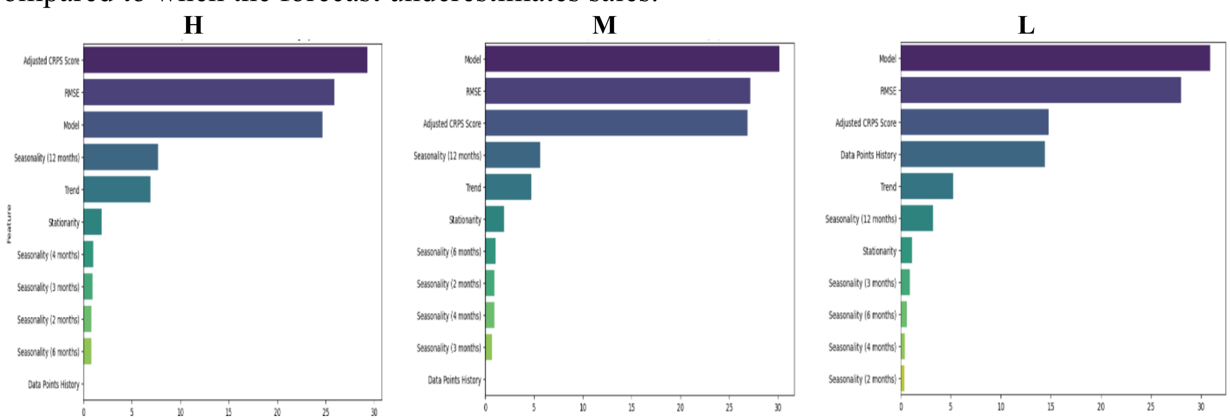

Figure 4. Feature importance in the CatBoost model for the three categories

Source: Author's contribution, generated using Python.

While the level of adjusted CRPS score is the most important feature that determines if a model is better performing in the class H of products, for the rest of the population of products, the categorical variables regarding the type of the model is the most relevant feature in identifying if an algorithm will better perform than the baseline model, as shown in Figure 4.

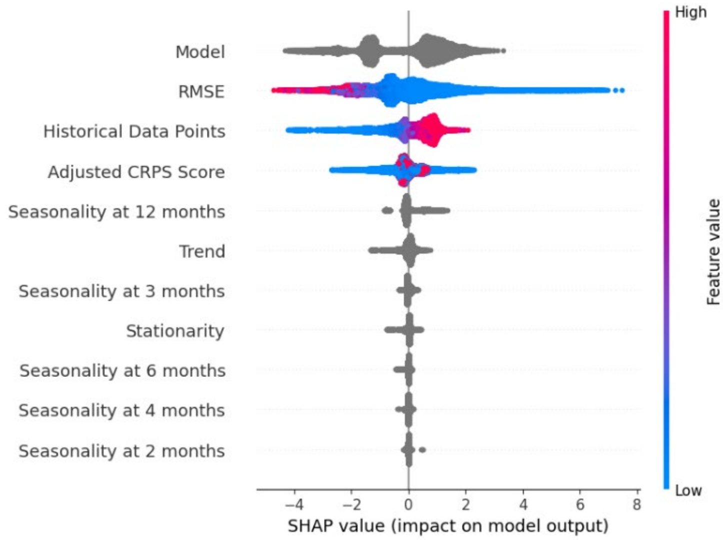

Figure 5. SHAP Feature impact on model output for category L

Source: Author's contribution, generated using Python.

If going deeper, into some of the predictors, for identifying their impact on the target, it is obvious that for distinct categories there are different outcomes. For example, as shown in Figure 5, the number of historical data points has a higher influence on determine a model to outperform the baseline model: the more values we have, the higher the probability of having an algorithm that outperforms the baseline model is.

The categorical variable: Model, is an important feature for all three clusters of products.

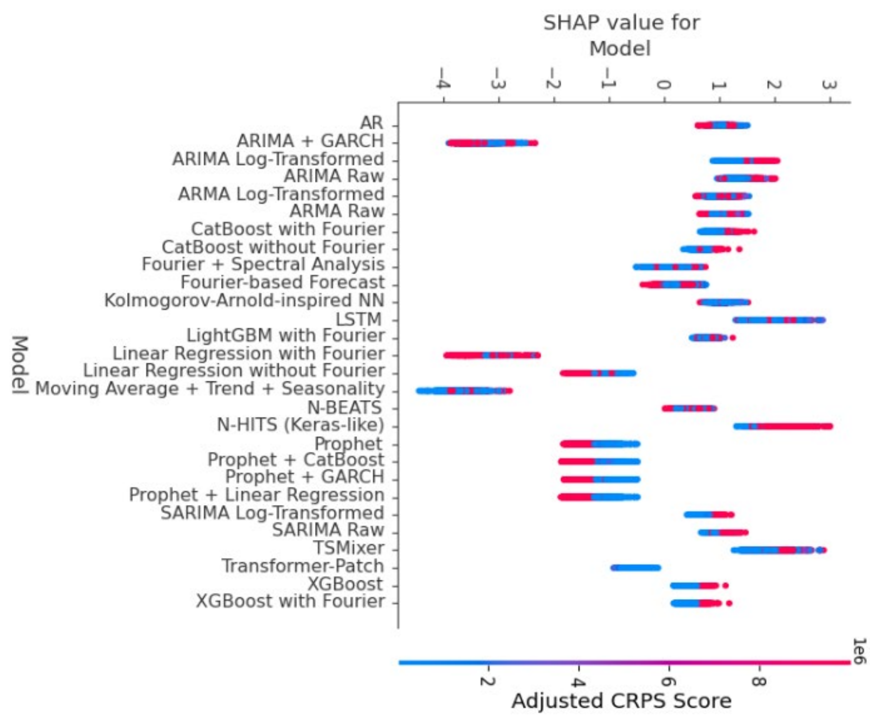

Figure 6. SHAP values for model and adjusted CRPS categories of predictors for category H

Source: Author's contribution, generated using Python.

DOI: 10.2478/picbe-2025-0103, pp. 1309-1322, ISSN 2558-9652 |
Proceedings of the 19th International Conference on Business Excellence 2025

In Figure 6, there is a detailed view of how models are distributed in terms of their potential impact on the outcome of the model: some models outperform the baseline model and models that tend to have a lower probability of doing so, as for example ARIMA with GARCH features, where ARIMA generates the forecast and the GARCH model captures ARIMA's errors and forecasts the volatility.

By testing different forecasting results from the trained algorithms, an issue arises. Some models were selected as the best in terms of the lowest RMSE, yet they failed to identify the most appropriate fluctuations in the data that would compare history with the future. As seen in Figure 7, the algorithm that provides the lowest RMSE predicts a constant evolution. If we use the CRPS score instead, this issue is eliminated because it evaluates the probability distribution of the entire forecast and compares it with historical dynamics, rather than assessing point-by-point errors as RMSE does.

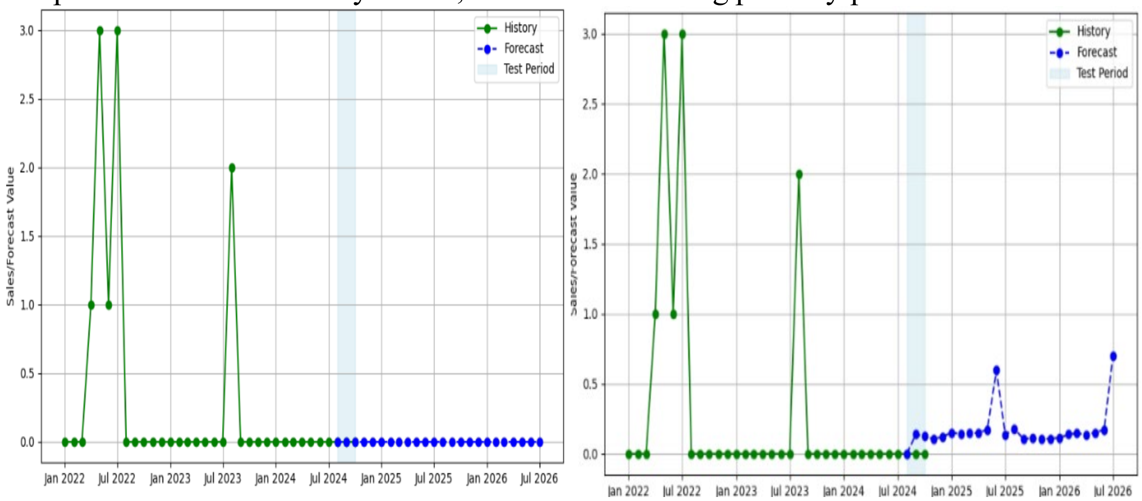

Figure 7. Difference in detecting the best performing model according to RMSE (left graph) and CRPS (right graph)

Source: Author's contribution, generated using Python.

The final step of the research was to optimize stock replenishment by calculating ROP, safety stock, and EOQ. To achieve this, a genetic algorithm was used with the objective of minimizing costs while ensuring product availability. The algorithm employs selection, crossover, and mutation to improve solutions over 100 generations, dynamically adapting to the monthly forecasted demand and lead times.

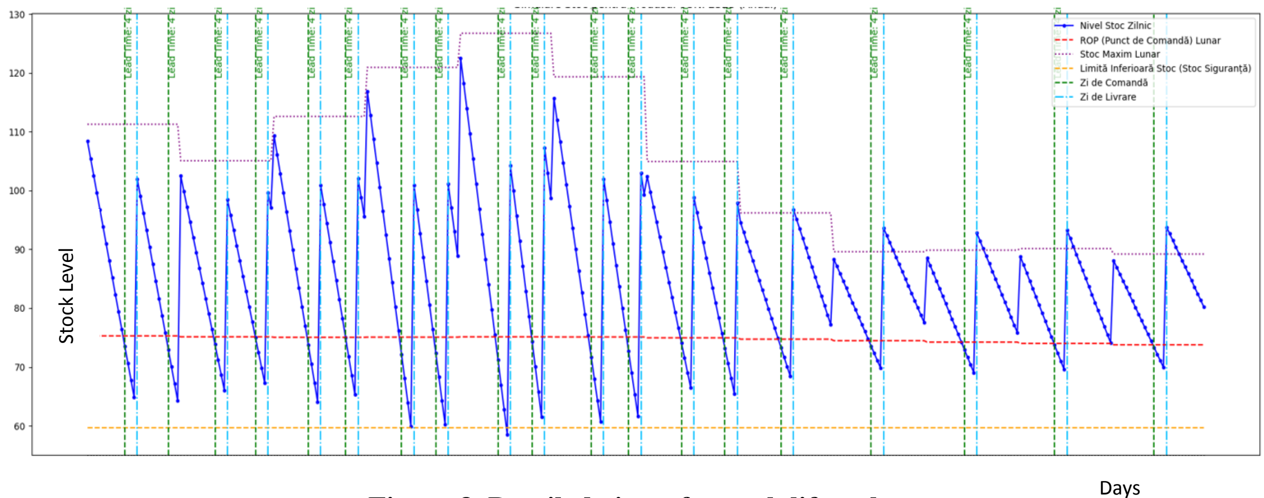

Figure 8. Detailed view of a stock lifecycle

Source: Author's contribution, generated using Python.

The objective was to be able to have a daily view of the stock flow over a longer period of time, as seen in the Figure 8, to be able to see the stock dilution, the point where the order needs to be put in place and the moment when the order is received and the stock increases with the reordered quantity.

Conclusions

The current research is a starting point for new and more detailed researches on forecasting, but also on identifying optimal methodologies for selection the best forecasting model and on developing and testing new optimization algorithms.

An important outcome of the current paper is the fact that it demonstrates the necessity of using more forecasting algorithms for detecting the future demand. Each timeseries has its own specificities and particularities, which may prevent certain algorithms accurately and completely capturing the information contained in the data. Some algorithms work well on normalized data, other work just fine on raw data, others detect better the trend and the seasonality, while others deal better with the noise in the data and can forecast better the moment when special sales will appear, like special sales detection, which can ne outliers or not.

As already identified in the case study, the more data we have, the better all the algorithms perform and in special time series cases, with a lower number of data points, the classical statistical methods outperform more complex algorithms. Essential in developing these models is a strong and deep understanding of the analyzed domain and the business requirements of that specific industry. Based on these elements, it becomes easier to decide which selection method to use in identifying the best-performing algorithm. If we use RMSE to select the best algorithm, we might choose a model that minimizes the error but does not correctly identify seasonality, trends, and time series dynamics. When the goal is to accurately capture the dynamics of the time series, including peak sales points, a better method is to use CRPS for selecting the best-performing model. However, even CRPS may not always be the best approach, as the objective might be to detect the overall dynamics without focusing on potential outliers or exceptional values. In such cases, it is necessary to build a hybrid score that combines multiple measurements in different proportions, determined through simulation or other algorithms.

From a practical perspective, retailers are advised to avoid applying a single forecasting model across all products. Instead, the results suggest using statistical models such as ARIMA or MA for products with limited sales history or irregular demand patterns. For products with more consistent sales and sufficient historical data, more complex models—such as Fourier-based methods or hybrid approaches—can significantly improve forecast accuracy. Boosting algorithms like LightGBM, especially when enriched with Fourier terms, are a good balance for medium-frequency products where patterns exist but are less stable. Retailers should also consider implementing a dynamic model selection process at the product level, using evaluation metrics like CRPS rather than relying solely on RMSE or MAPE, as this improves the alignment between model behavior and real-world demand fluctuations.

The optimization stage is essential in the cost reduction process and in increasing supply chain efficiency. It is crucial to have a deep understanding of the industry and the actual flows of that specific activity. It is also critical to know the essential cost parameters, lead time, and business requirements related to the cost function. In some industries, the objective is to never experience stock outages, while in others, a short stock outage is preferable to an overstock situation.

The limitations of this study primarily lie in the optimization phase, where only a single optimization algorithm—a genetic algorithm—was trained. Future work aims to test additional optimization algorithms to identify the most effective one or to develop algorithm architectures where multiple algorithms contribute to the final result. Another research direction will focus on identifying the optimal combination of forecasting methods, tailored to data characteristics and specific business requirements. Lastly, an additional research avenue in optimization will explore the ability of the optimization algorithm to adapt the forecasted quantity from either of two equally effective forecasting algorithms, depending on the optimization function and business needs.

Acknowledgments

This work was supported by the project "Societal and Economic Resilience within multi-hazards environment in Romania" funded by European Union – Nextgeneration EU and Romanian Government, under National Recovery and Resilience Plan for Romania, contract no.760050/ 23.05.2023

This paper was financed by the Bucharest University of Economic Studies during the PhD program.

References

Wang L.,Wnag X., Zhao Z. (2024). Mid-term electricity demand forecasting using improved multi-mode reconstruction and particle swarm-enhanced support vector regression. *Energy*, 304(132021).

Dalal S., Lilhore U.K., Simaiya S., Radulescu M.,Belascu L. (2024). Improving efficiency and sustainability via supply chain optimization through CNNs and BiLSTM. *Technological Forecasting and Social Change,* 209(123841).

Taillardat M., Fougères A.L., Naveau P., Fondeville R. (2023). Evaluating probabilistic forecasts of extremes using continuous ranked probability score distributions. *International Journal of Forecasting*, 39, 1448-1459.

Abolghasemi M., Tarr G., Bergmeir C. (2024). Machine learning applications in hierarchical time series forecasting: Investigating the impact of promotions. *International Journal of Forecasting,* 40, 597-615.

Ye L., Xie N., Boylan J.E., Shang Z. (2024). Forecasting seasonal demand for retail: A Fourier time-varying grey model. *International Journal of Forecasting*, 40, 1467-1485.

Corsini R.R., Costa A., Fichera S., Framinan J.M. (2024). Digital twin model with machine learning and optimization for resilient production–distribution systems under disruptions. *Computers & Industrial Engineering.* 191(110145).

Younespour M., Esmaelian M., Kianfar K. (2024). Optimizing the strategic and operational levels of demand-driven MRP using a hybrid GA-PSO algorithm. *Computers & Industrial Engineering.* 193(110306).

Wellens A.P., Boute R.N., Udenio M. (2024). Simplifying tree-based methods for retail sales forecasting with explanatory variables. *European Journal of Operational Research.* 314, 523-539.

Long X., Bui Q., Oktavian G., Schmidt D.F., Bergmeir C., Godahewa R., Lee S.P., Zhao K., Condylis P. (2025). Scalable probabilistic forecasting in retail with gradient boosted trees: A practitioner's approach. *International Journal of Production Economics.* 279(109449)

Wu Y., Meng X., Zhang J., He Y., Romo J.A., Dong Y., Lu D. (2024). Effective LSTMs with seasonal-trend decomposition and adaptive learning and niching-based backtracking search algorithm for time series forecasting. *Expert Systems with Applications*. 236(121202)

Ahmed S., Chakrabortty R.K., Essam D.L., Ding W. (2024). A switching based forecasting approach for forecasting sales data in supply chains. *Applied Soft Computing*. 167(112419)

Khedr A.M., S S.R. (2024). Enhancing supply chain management with deep learning and machine learning techniques: A review. *Journal of Open Innovation: Technology, Market, and Complexity*. 10(100379)

Sukolkit N., Arunyanart S., Apichottanakul A. (2024). An open innovative inventory management based demand forecasting approach for the steel industry. *Journal of Open Innovation: Technology, Market, and Complexity*. 10(100407)

Chae B., Sheu C., Park E.O. (2024). The value of data, machine learning, and deep learning in restaurant demand forecasting: Insights and lessons learned from a large restaurant chain. *Decision Support Systems*. 184(114291)

Keswani M. (2024). A comparative analysis of metaheuristic algorithms in interval-valued sustainable economic production quantity inventory models using center-radius optimization. *Decision Analytics Journal*. 12(100508)

Berrisch J., Ziel F. (2023). CRPS learning. *Journal of Econometrics*. 237(105221)

Koochali, A., Schichtel, P., Dengel, A., & Ahmed, S. (2022). Random Noise vs. State-of-the-Art Probabilistic Forecasting Methods: A Case Study on CRPS-Sum Discrimination Ability. *Applied Sciences*, *12*(10), 5104.

Appendix (if needed)