SCIENTIFIC OASIS

Spectrum of Decision Making and Applications

Journal homepage: www.dmap-journal.org
ISSN: 3042-1462

Retail Products Price Forecasting with Empirical Mode Decomposition and Auto Regressive Integrated Moving Average Model Using Web-Scraped Price Microdata

Mehmet Ozcalici[1,*], Elif Kaya[2]

[1] Department of International Trade and Logistics, Faculty of Economics and Administrative Sciences, Kilis 7 Aralik University, Kilis, Turkey
[2] Department of Economics and Finance, Faculty of Economics and Administrative Sciences, Istanbul Aydin University, Istanbul, Turkey

ARTICLE INFO

Article history:
Received 25 June 2024
Received in revised form 29 December 2024
Accepted 17 February 2025
Available online 23 February 2025

Keywords:
Retailing; Price forecasting; Web-scraping; ARIMA; Empirical Mode Decomposition (EMD).

ABSTRACT

This study presents a cutting-edge approach to price forecasting for an online retail business in Turkey, utilizing a hybrid model that combines Empirical Mode Decomposition (EMD) with Auto Regressive Integrated Moving Average (ARIMA) models. A 900-day dataset, scraped from the website, underpins this analysis. A battery of fourteen metrics is employed to evaluate the forecasting performance, culminating in a statistically significant confirmation of the hybrid model's superiority over the standalone ARIMA model, as established by the Wilcoxon signed-rank test. In addition to this performance validation, our investigation unveils an intriguing association between category standard deviations and forecasting accuracy, with lower standard deviations correlating with higher forecasting performance. While acknowledging the study's limitations related to data collection constraints, this research bears wider significance for the entire supply chain, offering strategic insights for retailers and the potential for more detailed analysis with larger datasets. Moreover, it lays the groundwork for future studies involving dynamic ARIMA parameter determination, advanced EMD variants, and machine learning integration, enhancing its applicability to various time series contexts. The results are compared with machine learning algorithms, namely Neural Networks, Support Vector Regression, Regression Tree, Gaussian Process Regression, and the Generalized Additive Model.

1. Introduction

Forecasting is one of the fundamental inputs to support planning decisions in retail chains [1]. In an era marked by the relentless evolution of e-commerce, understanding and forecasting retail price dynamics has become pivotal for both businesses and consumers alike. As online retail businesses proliferate, an intricate web of product categories emerges, each encapsulating a distinct set of market forces and consumer behaviors [2]. To grapple with the challenges posed by this dynamic

* *Corresponding author.*
E-mail address: mozcalici@gmail.com

https://doi.org/10.31181/sdmap21202519

landscape, this academic inquiry embarks on a rigorous exploration of retail price forecasting using advanced analytical tools.

The focal point of this study is the comprehensive analysis of retail category prices, hinging on the deployment of an ARIMA model. Going beyond traditional methods, we augment the ARIMA framework with EMD, a technique that dissects time series data into its constituent components, enabling a more nuanced understanding of price fluctuations. Our research scrutinizes the performance enhancement brought about by the inclusion of EMD in the modeling process.

The dataset underpinning our investigation encompasses the retail products of an online business, spanning 26 diverse categories, and spanning a time horizon of 893 days. This rich dataset serves as the foundation for the creation of 26 individual time series, each bearing unique insights into category-specific price dynamics. To ascertain the nonstationarity of these time series and their underlying components, we employ the Augmented Dickey-Fuller (ADF) test, a standard method in time series analysis.

In pursuit of greater forecasting accuracy, we implement EMD to deconstruct these time series into their constituent components, thereby laying the groundwork for an augmented ARIMA-based modeling approach. This augmented model is meticulously designed to forecast future values for each component and residual, subsequently integrating these forecasts to yield the final output values.

The optimization of the ARIMA model's parameters is a crucial step in our methodology. We achieve this by carefully examining Bayesian Information Criterion (BIC) values, ensuring that our model is fine-tuned to yield the most reliable forecasts.

Our evaluation of model performance extends to 14 different metrics, offering a comprehensive view of the models' efficacy in forecasting the future average prices for each category. The results unequivocally demonstrate that the integration of EMD significantly bolsters the forecasting performance of the single ARIMA model, lending statistical significance to this augmentation.

Beyond this, our inquiry ventures into the relationship between the standard deviation of time series data and its impact on forecasting performance. To this end, we employ linear regression models, with the standard deviation of mean category prices as the independent variable and the performance evaluation metrics as the dependent variable. The results reveal a pronounced trend, with forecasting performance markedly superior in categories where the standard deviation is smaller, underscoring the pivotal role of price stability in forecasting accuracy.

In a landscape where data-driven insights are the linchpin of effective decision-making, this study illuminates a pathway towards enhanced retail price forecasting. By juxtaposing ARIMA modeling with EMD and shedding light on the importance of standard deviation, we endeavor to provide a substantial contribution to the arsenal of tools available to researchers and practitioners navigating the complex world of online retail.

With the entry of additional large players, increasing multichannel activities on the supply side, and the emergence of new store formats and market channels, retailers are facing heterogeneous and changing consumer preferences [3]. As a result, they are increasingly competing in both price and variety.

Forecasting future values represents a pivotal domain within the realm of data science which involves building a model using past observations as a basis for forecasting future values [4]. The autoregressive integrated moving average (ARIMA) has been used in several fields for statistical analysis and data prediction, such as forecasting copper spot price [5], carbon price [6–8], coal price [9,10], metal price [11], electricity price [12–14], oil price [10,15,16], natural gas price [10], fuelwood price [17], stock price or index [18–22], cryptocurrency price [19,23], exchange rate [24], inflation of various countries [25,26], timber price [27], agricultural future price [12,28,29], computer RAM price

[30], gold price [31], coffee price [32], uranium price [33], salmon price [34], rice price [35], algorithmic investment strategies [36], and so on. The one thing these studies have in common is to apply the ARIMA technique to successfully forecast the various time series that have an impact on people's lives.

In retailing industry demand [37–40], sales [41–44], timing of retail orders [45] and market share [46] forecasting is successfully applied. In a regression-based ML study [47] for retail price optimization leveraging product, competitor, and customer factors, Random Forest (94% accuracy) outperforms Decision Tree and Logistic Regression, enabling retailers to refine pricing strategies for maximum profit and competitiveness. Forecasting on the retail prices of various products which is also a time series are not studied in depth before. One of the main reasons for this could be the cost associated with tracking prices daily in the past. Nowadays, it has become possible to record retail product prices on a daily basis through the internet via automated systems such as web scraping. It is, thus, possible to collect and analyze a large amount of data.

Although the ARIMA-based models are mainly suitable for linear time-series analysis [48], its performance can be further be improved by integrating with Empirical Mode Decomposition (EMD) technique [49–52].

A hybrid EMD-ARIMA method for time series forecasting is presented in this paper. The available data is treated as a time series and key features of past data are identified to forecast future values of the sequence using our approach. The advantages of decomposing time series with EMD and forecasting each component with ARIMA and then integrating the results to obtain final forecasting value includes, increasing performance in forecasting, a reliable financial time series forecasting tool and it does not require prior assumptions of the empirical data features since the model is derived from empirical data. In this study, the relationship between the standard deviation values of time series and their forecasting performance is also examined. As a result, as expected, higher forecasting performance is determined in time series with lower standard deviation values.

Our contribution to the literature can be presented as follows: (1) Scraping web-based price data (similar to used in [53], (2) applying ARIMA method to the dataset to obtain forecasted prices, (3) apply EMD to dataset and for each IMF and residual apply ARIMA method and finally integrate the results to obtain more powerful price forecast, (4) analyze the relationship between standard deviation and forecasting performance of categories.

The research questions to be answered in the study are as listed below:

- Comparing the performance of ARIMA techniques with and without using EMD signal decomposition.
- Analyze the relationship between standard deviation and forecasting performance.

The remainder of the paper is organized as follows: following this introductory chapter, a review of relevant research is presented in chapter two. The third section is devoted to the basic principles of the methodology. The fourth section is devoted to the description of the dataset and the results of the analyses. Finally, the fifth section contains the concluding remarks.

2. Literature review

In today's dynamic e-commerce environment, where accurate forecasting of retail prices is of paramount importance, a thorough understanding of the existing body of knowledge is essential to make informed contributions.

This section begins by reviewing the key studies that have laid the groundwork for price forecasting models. It then discusses recent developments and emerging trends in the field, covering both traditional methodologies and the latest approaches used in various studies. Through this research, we aim to build a solid theoretical framework underpinning our own innovative hybrid model combining Empirical Mode Decomposition (EMD) and Autoregressive Integrated Moving Average (ARIMA) techniques.

Claveria and their colleagues used survey expectations about a wide range of economic variables to forecast real activity by using evolutionary algorithms [54]. In their study, scatter diagrams where the x axis shows the standard deviation of GDP growth while the y-axis shows the forecast accuracy measures are presented. It shows that the scatter diagrams have positive slope which indicates that standard deviation and forecasting performance (smaller values) is positively related.

Liu and colleagues employed EMD-recursive ARIMA method to develop a wind warning system to protect the running trains under strong crosswind along the Chine Qinghai-Tibet railway [49]. The results indicate that their proposed hybrid method has satisfactory performance when compared with single ARIMA model.

Liu and colleagues employed EMD with various forecasting methods including ARIMA to predict the wind speed [55]. At the end of the study, they also report that hybrid models are increasing the performance of single models. Moreover, they report that ARIMA can effectively predict the low-frequency sub-sequences after EMD.

In their study, Cao and colleagues [56] presents a hybrid approach that integrates an extended ARIMA model, Empirical Mode Decomposition (EMD), and truncated Singular Value Decomposition (SVD) to predict multiple air quality indicators across several monitoring stations. Experimental results show that the proposed model surpasses state-of-the-art forecasting methods in both accuracy and time efficiency

Wang and colleagues are employed EMD-ARIMA method to predict the traffic speed [57]. They tested the proposed method on four distinct scenarios. They evaluated the proposed hybrid method against the experimental data and compared with the results from the traditional models. As a result, they report that hybrid proposed method outperforms the traditional forecasting models.

Abadan and Shabri applied hybrid EMD-ARIMA model to forecast the prices of rice in Malaysian Ringgit per metric ton [58]. It is reported that EMD increased the accuracy of the ARIMA forecasting.

Nasir and colleagues [59] proposes a novel hybrid LMD-SD-ARIMA-LSTM model to predict crude oil prices by decomposing the original data into stochastic and deterministic components, then integrating ARIMA and LSTM to capture both linear and nonlinear trends. The results show that this method achieves superior performance (lowest MAE and MAPE) in both short- and long-term forecasting compared to existing models, providing a valuable tool for informed decision-making in the global energy sector.

Li and colleagues are applied the hybrid EMD-ARIMA method to predict the Shenzhen Growth Enterprise Market Price Index [51]. They reconstructed the IMFS to obtain high frequency, low frequency and residual sequences. Thus, they clustered the IMF and residual components to three clusters. Which enables a less complex variable set, in which they report that the hybrid method increased the prediction performance of the single ARIMA model.

Wang and colleagues are employed the hybrid EMD-ARIMA and EEMD-ARIMA models for long-term streamflow forecasting[50]. They compared the performance of the models and report that the best performing model is the EMD-ARIMA hybrid model, among the other models.

Nuzhat and colleagues are employed the hybrid EMD-ARIMA method to forecast the medical tourism [52]. They report that proposed hybrid forecasting approach has outperforming characteristics.

Yang and Lin combined EMD, ARIMA and SVR to forecast the stock indices [60]. In their proposed approach, the ARIMA model is employed to analyze the linear part of the original time series and EMD is used to decompose the dynamics of the non-linear part of the dataset. At the end of the paper, it is reported that proposed approach is superior to other benchmark models.

Zhou and Huang applied the hybrid EMD-ARIMA model to predict the remaining useful life of lithium-ion batteries [61]. At the end of the study, it is reported that the proposed hybrid EMD-ARIMA method outperformed other benchmark models.

Büyükşahin and Ertekin applied the hybrid EMD-ARIMA model with various variations to four different datasets including a financial time series namely GPD/USD [24]. At the end of the study, they report that, forecasting performance gets better if more stationary time series data is provided.

Awajan and colleagues [62], proposed an enhanced EMD-ARIMA hybrid approach to forecast daily COVID-19 cases in Jordan by decomposing the data into IMFs and applying ARIMA to the low-frequency components, ultimately outperforming other forecasting methods in predictive accuracy.

In summary, these studies consistently report enhanced forecasting performance when employing hybrid models, with EMD-ARIMA often outperforming single models and benchmark approaches. Furthermore, the literature underscores the adaptability of hybrid models across diverse datasets and applications. The collective findings from these studies lay a robust foundation for our own research, which introduces a novel EMD-ARIMA hybrid model for price forecasting in the context of online retail. Building on the successes and insights of prior research, we aim to contribute to this growing body of knowledge by evaluating the performance of our hybrid model in the unique setting of retail category price forecastings.

3. Methodology

3.1 Auto-regressive integrated moving average (ARIMA)

ARIMA, which stands for Auto Regressive Integrated Moving Average, is a popular statistical model used for analyzing and forecasting time series data. It is often considered one of the best models for time series data because of its ability to capture complex patterns and trends in the data [63]. The ARIMA model forecasts an assumed variable's future value with several past observations and random errors with a linear function [64].

The ARIMA model consists of three main components: the AR model, the MA model, and the integration component. The AR model, or Auto Regressive model, is a time series model that uses past values of the variable being analyzed to forecast future values. The MA model, or Moving Average model, is a time series model that uses the error term from past predictions to forecast future values. The integration component is used to transform the data so that it becomes stationary, which is necessary for the AR and MA models to be effective.

The ARIMA (p,d,q) model has three main parameters that are used to control the behavior of the model. The first parameter is the autoregressive lags, denoted by p. This parameter specifies the number of lagged values of the variable being analyzed that will be included in the model. The parameter "p" in the autoregressive (AR) model can be ascertained through the analysis of the Partial Auto-correlation Function (PACF) plot, which identifies the correlation between residuals and their lagged counterparts. The determination of the order of the AR term "p" is contingent upon the count of non-zero partial autocorrelations within the PACF plot.[65].

The second parameter is the moving average, denoted by q. This parameter specifies the number of lagged errors that will be included in the model. Autocorrelation Function (ACF) plot is used to identify the MA term "q". The Autocorrelation Function (ACF) plot elucidates the degree of association between the current observation in a time series and its preceding values. The lag at which the ACF ceases to exhibit significant non-zero autocorrelations signifies the order of the Moving Average (MA) term denoted as "q." [66].

These two parameters (p and q) can also be determined by AIC or BIC criteria. The ARIMA (p, d, q) model with a lower AIC or BIC value is a better model [67].

Finally, the third parameter is the order of differentiation, denoted by d. This parameter specifies the number of times the data will be differenced to achieve stationarity. The first difference is defined as

$$y'_t = y_t - y_{t-1} \tag{1}$$

The second difference is defined as;

$$y''_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2} \tag{2}$$

Three parameters are shown as ARIMA (p, d, q) to represent the model. An ARIMA model can be expressed as in the equation;

$$y'_t = c + \varphi_1 y'_{t-1} + \cdots + \varphi_p y'_{t-p} + \vartheta_1 \varepsilon_{t-1} + \cdots + \vartheta_p \varepsilon_{t-p} + \varepsilon_t \tag{3}$$

where y' is a stationary time series (which can be the d-time differenced series of the original non-stationary series - y), φ_t are the parameters of the autoregressive terms, ϑ_t are the parameters of the moving average terms and ε_t are the error terms.

This study aimed to determine the optimal parameters for the ARIMA model, specifically the “d” parameter, which determines the order of differentiation, using the ADF test. The ADF test was run for three different models (a model without trend and constant, a model with constant, and finally a model with both constant and trend). If the series was found to not be stationary in any of these models, the next “d” value is tested until the series became stationary. The “p” and “q” values were also tested up to a value of 5, and the BIC value is calculated for each combination of “p” and “q” values. This resulted in a matrix of BIC values with dimensions of 5x5. The “p” and “q” values with the lowest BIC value were determined as optimal for the ARIMA model.

3.2 Empirical mode decomposition (EMD)

The EMD algorithm was introduced by Huang and his research associates [68]. This method involves the decomposition of intricate original signals into a set of Intrinsic Mode Functions (IMFs) characterized by varying amplitudes, along with a residual component. At its core, EMD facilitates the transformation of nonlinear and non-stationary signals into linear and stationary forms [69].

Each IMF must satisfy two conditions [70]: a) Within the entire dataset, it is imperative that the count of extrema and the count of zero-crossings exhibit parity, with a permissible variance not exceeding one b) The average value of the upper and lower envelopes, derived from the signal's maximal and minimal values, is constrained to zero, signifying that these envelopes exhibit local symmetry relative to the zero axis.

The computational steps of the EMD are given as follows [68]:

Step 1: Determine all the local extrema of the series $\{\mathcal{S}(t)\}$, encompassing both local maxima and local minima.

Step 2: Establish a cubic spline curve by connecting all the local maxima to create the upper envelope denoted as $\{\mathcal{S}_{up}(t)\}$. Analogously, the lower envelope, denoted as $\{\mathcal{S}_{low}(t)$, is constructed using all the local minima.

Step 3: Determine the mean envelope $\{\mathcal{M}(t)\}$ by calculating it from the upper and lower envelopes using the following equation:

$$\mathcal{M}(t) = \frac{\mathcal{S}_{up}(t) + \mathcal{S}_{low}(t)}{2} \tag{4}$$

Step 4: Extract the details using Equation 5.

$$\mathcal{Z}(t) = \mathcal{S}(t) - \mathcal{M}(t) \tag{5}$$

Step 5: Check whether $\{\mathcal{Z}(t)\}$ is an IMF: (a) if $\{\mathcal{Z}(t)\}$ is an IMF then set $\mathcal{C}(t) = \mathcal{Z}(t)$ and meantime replace $\{\mathcal{S}(t)\}$ with the residual $\mathcal{R}(t) = \mathcal{S}(t) - \mathcal{C}(t)$; (b) if $\{\mathcal{Z}(t)\}$ is not an IMF, replace $\{\mathcal{S}(t)\}$ with $\{\mathcal{Z}(t)\}$ then repeat Steps 2–4 until the termination criterion is satisfied. The following equation can be regarded as the termination condition of this iterative calculation:

$$\sum_{t=1}^{m} \frac{\left[\mathcal{Z}_{j-1}(t) - \mathcal{Z}_{j}(t)\right]^2}{\left[\mathcal{Z}_{j-1}(t)\right]^2} \leq \epsilon \; (j = 1,2, \dots ; t = 1,2, \dots , m) \tag{6}$$

where m is the number of the wind speed data points, ϵ is the terminated parameter, and 'j' denotes the times of iterative calculation. In this study, the ϵ is set equal to 0.2.

Finally, n orthogonal IMFs are reconstructed by summation.

$$\mathcal{S}(t) = \sum_{i} IMF_i(t) + \mathcal{R}(t) \tag{7}$$

Step 6: The procedure of Steps 1–5 is repeated until all the IMFs are found.

Nonetheless, the Empirical Mode Decomposition (EMD) method is associated with a significant issue, as pointed out by Zhang and colleagues, which is commonly known as modal aliasing [70]. Modal aliasing refers to a situation in which signals characterized by distinct scales or frequencies are erroneously contained within the same Intrinsic Mode Function (IMF) component, or when signals with similar scales or frequencies are erroneously distributed across multiple disparate IMF components.

3.3 White noise test

The Ljung-Box test constitutes a statistical methodology primarily employed for assessing the presence of white noise within a given data series. In the context of this test, the null hypothesis posits that the data series is characterized by independent distribution, thus conforming to the attributes of white noise. In contrast, the alternative hypothesis challenges this assumption by proposing that the data series deviates from independent distribution and instead displays evidence of serial correlation, a characteristic suggesting temporal dependencies [71].

It is important to note that in the realm of time series analysis, the identification of white noise is a critical consideration. It serves as a fundamental prerequisite for the effective modeling of time series data. In essence, time series data can only be reasonably and accurately modeled when it exhibits departures from white noise characteristics, as the presence of temporal dependencies and serial correlation typically underlie the underlying dynamics of many real-world time series datasets [71].

3.4 Performance metrics pool

In academic literature, a definitive and universally accepted criterion for evaluating models lacks consensus [72]. Number of performance evaluation metrics are developed. Studies on forecasting financial time series have been reviewed, and fourteen metrics commonly used in these studies have been compiled. When the first four of these metrics take on higher values, the forecasting performance is better, while the remaining ten metrics perform better when they take on larger values. The metrics and their formulas are listed in Table 1.

Table 1
Performance evaluation metrics pool

Characteristics	Name of the Metric	Equation of the Metric
Higher is Better	R^2	$1-\frac{\sum_{i=1}^{n}(\hat{p}_i-p_i)^2}{\sum_{i=1}^{n}(p_i-\bar{p})^2}$
	Adjusted R^2	$1-\frac{(1-R^2)(n-1)}{(n-w-1)}$
	Hit Ratio (HR)	$HR=\frac{number\ of\ hits}{n}$
	Directional Symmetry (DS)	$DS=\frac{1}{n}\sum_{(i=1)}^{n}\mu_i$ $\mu_i=\begin{cases}1, & (p_{t+1}-p_t)(\hat{p}_{t+1}-\hat{p}_t)>0\\ 0, & others\end{cases}$
Smaller is Better	Mean Absolute Error (MAE)	$\frac{1}{n}\sum_{i=1}^{n}\lvert d_i\rvert$
	Mean Squared Error (MSE)	$\frac{1}{n}\sum_{i=1}^{n}d_i^2$
	Square-root of MSE (RMSE)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}d_i^2}$
	Mean Absolute Relative Error (MARE)	$\frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{d_i}{p_i}\right\rvert$
	Mean Square of Relative Error (MSRE)	$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{d_i}{p_i}\right)^2$
	Square-root of MSRE (RMSRE)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{d_i}{p_i}\right)^2}$
	Mean Absolute Percentage Error (MAPE)	$\frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{d_i}{p_i}\right\rvert\times 100$
	Mean of Square Percentage Error (MSPE)	$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{d_i}{p_i}\right)^2\times 100$
	Square-root of MSPE (RMSPE)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{d_i}{p_i}\right)^2\times 100}$
	U-Theil	$\frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}d_i^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}p_i^2}+\sqrt{\frac{1}{n}\sum_{i=1}^{n}\hat{p}_i^2}}$

$p; d_i=p_i-\hat{p}_i$
p_i: actual price for day i
$\hat{p}_i$: predicted price for day i
n: number of days
$number\ of\ hits$: correctly predicted price increase or decrease
w : number of independent variable.

3.5 Proposed methodology

The outline of the study is presented in Figure 1.

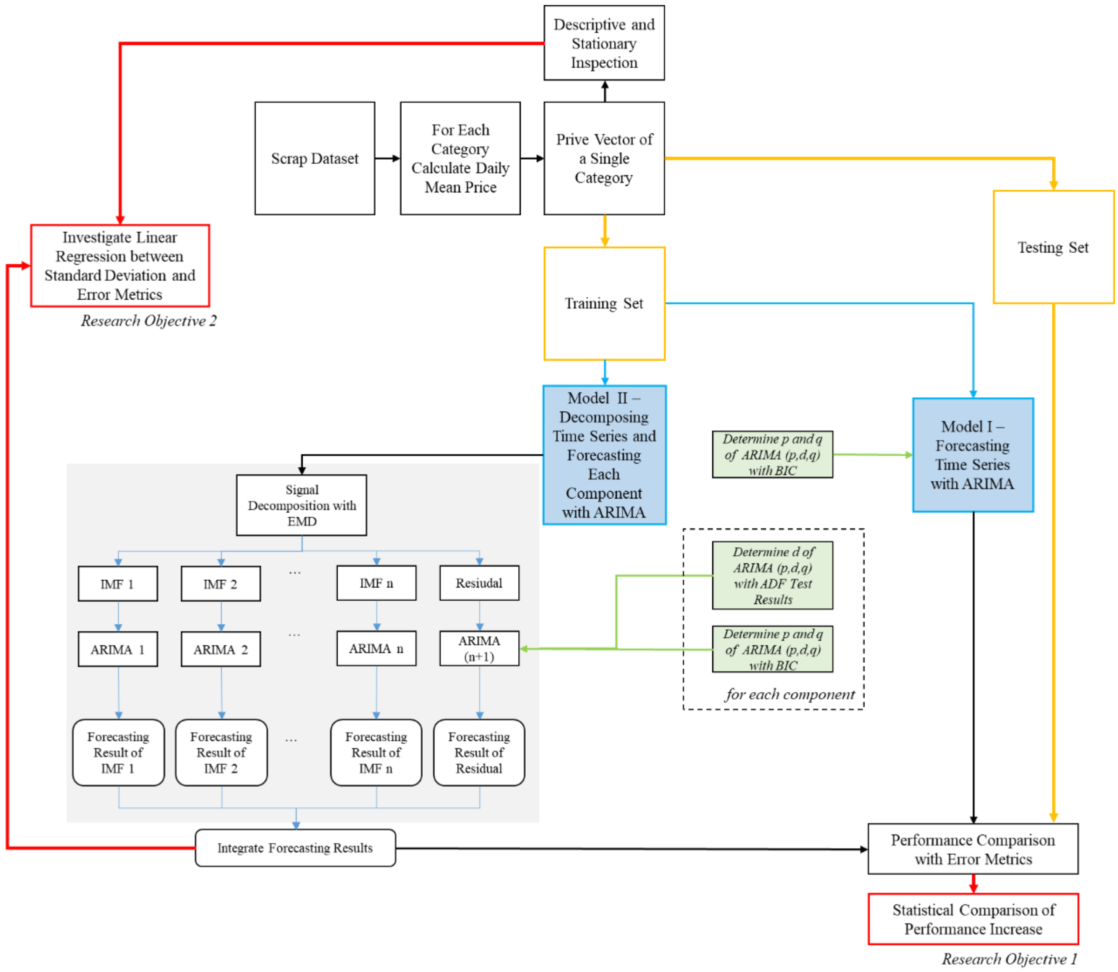


Fig. 1. Flowchart of the study

The steps of the study can be listed as follows:

i. First, the prices and names of the products were scraped on the website every day for eight hundred ninety-three days.
ii. The arithmetic average of the prices of the products on each day was calculated for each category. Thus, the time series was obtained as much as the number of categories.
iii. The profiles of the time series are revealed with the help of descriptive statistics. In addition, whether the time series are stationary or not was examined with the help of the Augmented Dickey Fuller test.
iv. Each time series is divided into two subsets. The last ninety observations were set aside to test the performance of the models, and with the training set containing the remaining 803 days of data, the parameters of the models were determined, and the models were trained.
v. The performance of two models was examined using the training set using two models:
vi. Model 1. In this model, the next steps of the time series are estimated with the ARIMA model. The parameters of the ARIMA model were determined with the help of BIC. Using the forecasting scheme, forecasts were made for each day in the training set and the

results were compared with actual prices and forecast performance metrics in the pool were calculated.

vii. Model 2. In this model, each time series is first separated into its components. EMD algorithm is used for this decomposition. Each category is divided into a different number of IMF and residual components. Then, for each component, the ARIMA model was run under the estimation scheme. The best p, d and q values for the components differ from each other in each category. Forecasts were collected for each component, thus forecasting the value of the category for the next day. After this stage, just as in Model 1, the actual values and the estimated values of Model 2 were evaluated by calculating the metrics in the pool.

viii. In accordance with the first research question, the forecasting performances of Model 1 and Model 2 were compared. Wilcoxon signed rank test was performed to test the differences.

ix. Finally, the relationship between the standard deviation values of the categories and the estimation performance values was examined.

Forecasting scheme is presented in Figure 2. The models are trained with the training set using optimized ARIMA parameters, and forecasting are made for the next day. Every day, new prices are added to the training vector, and thus, the number of data points in the training set increases over time. For example, the model is trained with a training vector containing 803 days for the first day of the test set, and a forecasting is made for the next day. For the last day of the test set, the model is trained with a training vector containing 892 days, and a forecasting is made for the last day.

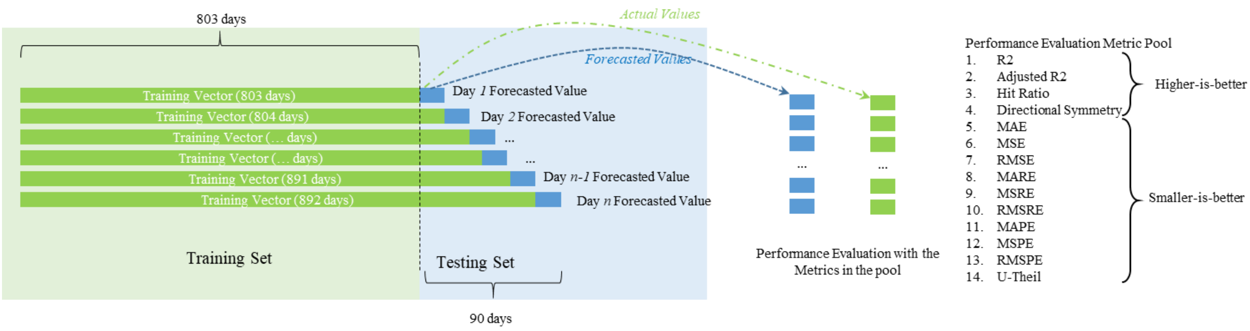


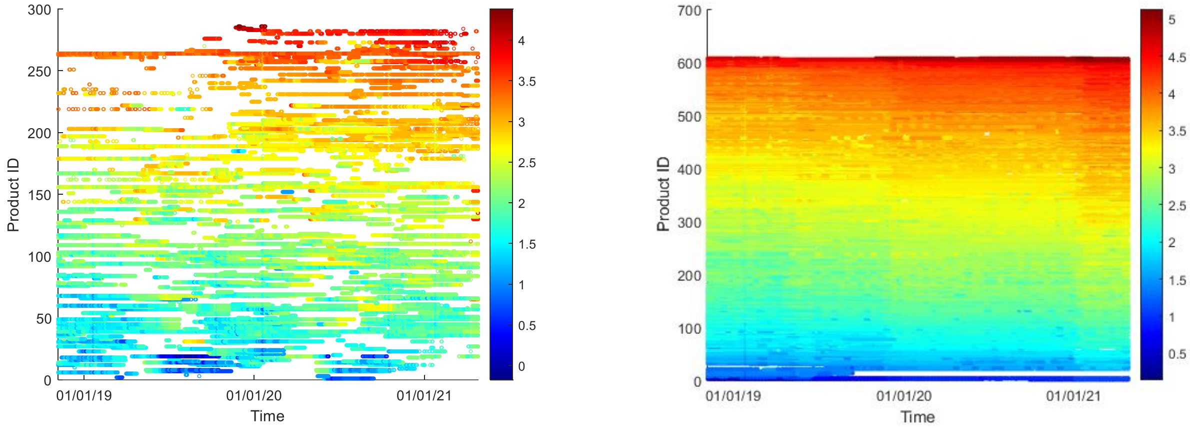
Fig. 2. Forecasting scheme

4. Analysis

4.1 Dataset description

The dataset used in this study was scraped from one of the prominent shopping mall chains in Turkey that allows customers to buy groceries online. This website was a pioneer in this area, and as a result, other retail chains also started providing the option of online shopping. The data collection took place between November 6th, 2018, and April 26th, 2021, which covers a period of 893 days. The data was scraped for a total of 26 categories.

Plots of the web-scraped data for two product categories namely fruit and cheese are presented in Figure 3. In figure the extent of the data missingness can be seen from the white gaps in the arrays. There are more missing values in the fruit category when compared with the cheese category. One of the reasons for this is that diverse fruits are made available for sale during various seasons. However, it is evident that a consistent number of products are introduced to the sale within the cheese category throughout the year.



(a) (b)

Fig. 3. Web-scraped daily log-price data for products over time (a) Fruit Category (b) Cheese Category

During the period examined within the scope of this study, the nation's economy experienced a phase of comparatively modest inflation rates. However, subsequent to the cessation of the data scraping period, disruptions transpired within the economic indicators, precipitating an escalation in the inflation rate and a depreciation of the Turkish lira vis-à-vis the US dollar. The deliberate decision to develop prediction models during a period of economic stability, coupled with the subsequent assessment of these models under conditions of elevated inflation, inherently engenders suboptimal predictive performance. Hence, the study was methodically executed, utilizing the dataset procured during the aforementioned interval to ensure the integrity and reliability of the predictive models' performance evaluations.

After extracting price data for products in the categories, the average price was calculated for each category. This average price is the arithmetic mean of the prices of the products offered for sale in the category on that day. A box plot of the average price for each category is presented in Figure 4. The first 21 categories consist of food and beverage products, while the last five categories consist of cleaning products.

To examine the forecasting performance, the observations containing the last 90 days (January 19, 2021 to April 26, 2021) were set aside. Thus, while the price information for 803 days (November 6, 2018 to January 18, 2021) was used in developing the models, the last 90 days of observations were used to examine the out-of-sample performance of the developed models.

Descriptive information about the categories for the training set is presented in Table 2. It is possible to create a similar table for the out-of-sample dataset as well. The table lists the number of unique products listed in each category, as well as the minimum, maximum, average, median price, and standard deviation of the price. In addition, the table includes the skewness and kurtosis values, as well as the JB statistics to test if the price distribution is normal. The Bakery and Patisserie category has the highest number of product types with 1,289 products, while the Sugar category has the lowest number of product types with only 48 products. The p-value from the JB test for products other than shaving supplies being lower than the 0.01 level indicates that the prices of these products are not normally distributed. The positive skewness values indicate that the right side of the distribution of the mean price is longer, meaning that there are few products with prices higher than the mean.

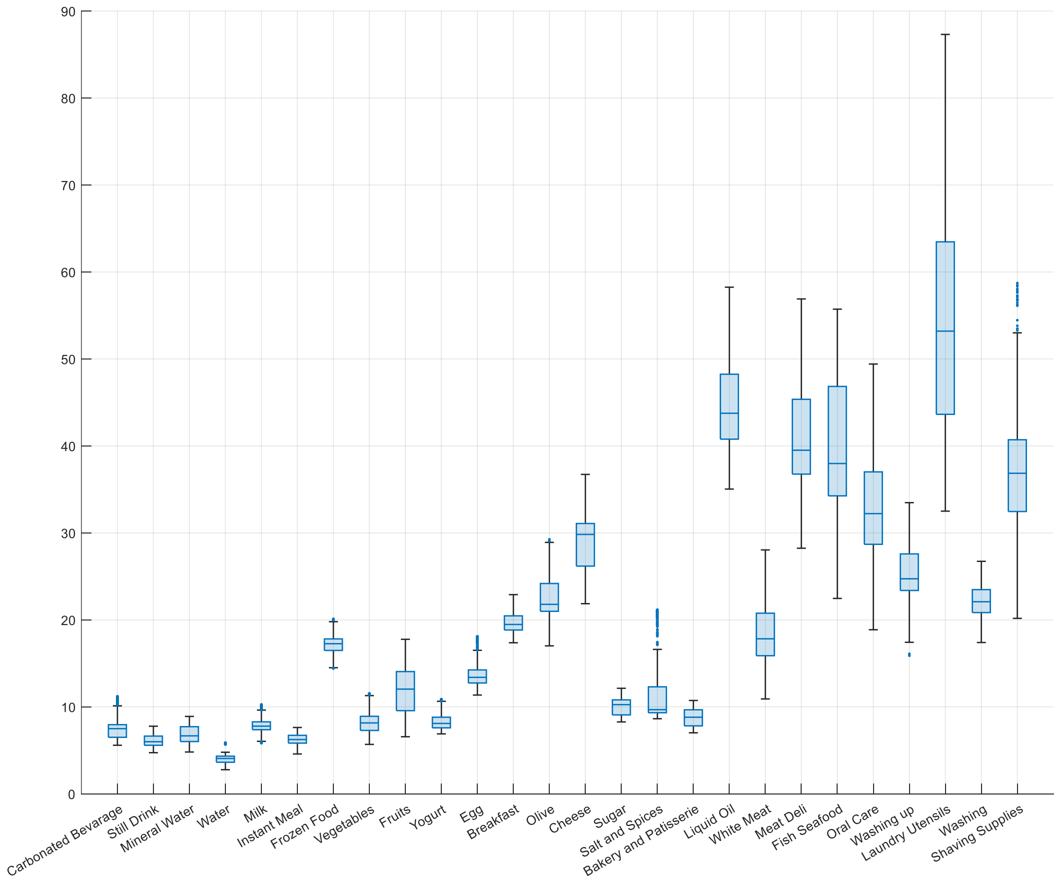


Fig. 4. Box plot of mean price for each category

Table 2
Descriptive statistics for training set

Categories	Number of SKUs	Min	Max	Mean	%25	Median	%75	Standard Deviation	Skew-ness	Kurt-osis	JB Statistics*
Carbonated Beverage	443	5.59	11.21	7.49	6.47	7.44	7.91	1.38	0.89	3	112.45
Still Drink	789	4.73	7.78	6.05	5.57	5.94	6.58	0.66	0.25	2.42	20.78
Mineral Water	171	4.81	8.83	6.77	6	6.62	7.67	0.98	0.23	1.93	48.16
Water	169	2.78	5.87	3.96	3.6	4.04	4.28	0.57	0.65	4.74	137.04
Milk	363	5.83	10.11	7.81	7.35	7.76	8.18	0.77	0.42	3.54	31.01
Instant Meal	152	4.58	7.63	6.16	5.78	6.23	6.57	0.71	-0.37	2.47	29.55
Frozen Food	238	14.42	20.12	17.07	16.44	17.26	17.82	1.19	-0.3	2.76	14.89
Vegetables	440	5.68	10.87	8.04	7.26	8.07	8.75	1.07	0.21	2.45	16.82
Fruits	286	6.57	17.77	11.66	9.5	11.72	13.85	2.5	0.09	1.76	55.7
Yogurt	318	6.89	10.89	8.21	7.55	8.05	8.71	0.88	1.12	4.27	233.58
Egg	68	11.36	17.87	13.6	12.72	13.34	13.97	1.29	1.31	4.62	280.32
Breakfast	625	17.37	22.87	19.72	18.82	19.37	20.23	1.17	0.87	2.86	105.28
Olive	181	17.01	29.27	22.46	20.97	21.71	24.05	2.36	0.63	3.29	59.63
Cheese	607	21.86	36.72	28.66	26.06	29.57	30.94	3.11	0.21	2.27	25.11

Table 2
Continued

Categories	Number of SKUs	Min	Max	Mean	%25	Median	%75	Standard Deviation	Skew-ness	Kurt-osis	JB Statistics*
Sugar	48	8.27	12.14	10.01	9.03	10.17	10.76	0.89	-0.3	1.77	66.64
Salt and Spices	372	8.64	21.17	11.75	9.32	9.64	12.49	4.15	1.45	3.36	206.33
Bakery and Patisserie	1289	7.02	10.64	8.72	7.8	8.61	9.57	0.9	0.1	1.62	69.49
Liquid Oil	230	35.04	56.02	44.62	40.69	43.55	47.89	4.81	0.65	2.4	72.13
White Meat	266	10.91	28.04	18.11	15.81	17.63	20.26	3.11	0.59	2.9	49.06
Meat Deli	438	28.24	53.6	40.6	36.63	39.07	44.15	5.75	0.4	2.31	39.67
Fish Seafood	134	22.47	55.32	39.63	34.09	37.7	46.02	6.95	0.43	1.91	67.87
Oral Care	685	18.87	45.48	32.7	28.57	31.79	36.21	5.39	0.46	2.51	37.79
Washing up	232	16.11	33.48	25.14	23.31	24.62	27.19	3.16	0.36	3.09	18.42
Laundry Utensils	98	32.5	87.31	54.03	43.33	50.66	62.87	13.42	0.41	2.32	40.43
Washing	804	17.4	26.21	22.13	20.81	21.95	23.19	1.65	0.29	2.63	16.9
Shaving Supplies	351	20.17	53.47	36.29	32.38	36.47	40	6.4	0.01	2.98	0.04

* All of the statistics are significant at 0.01 level except for the Category 26: Shaving Supplies

4.2 Model 1 forecasting results

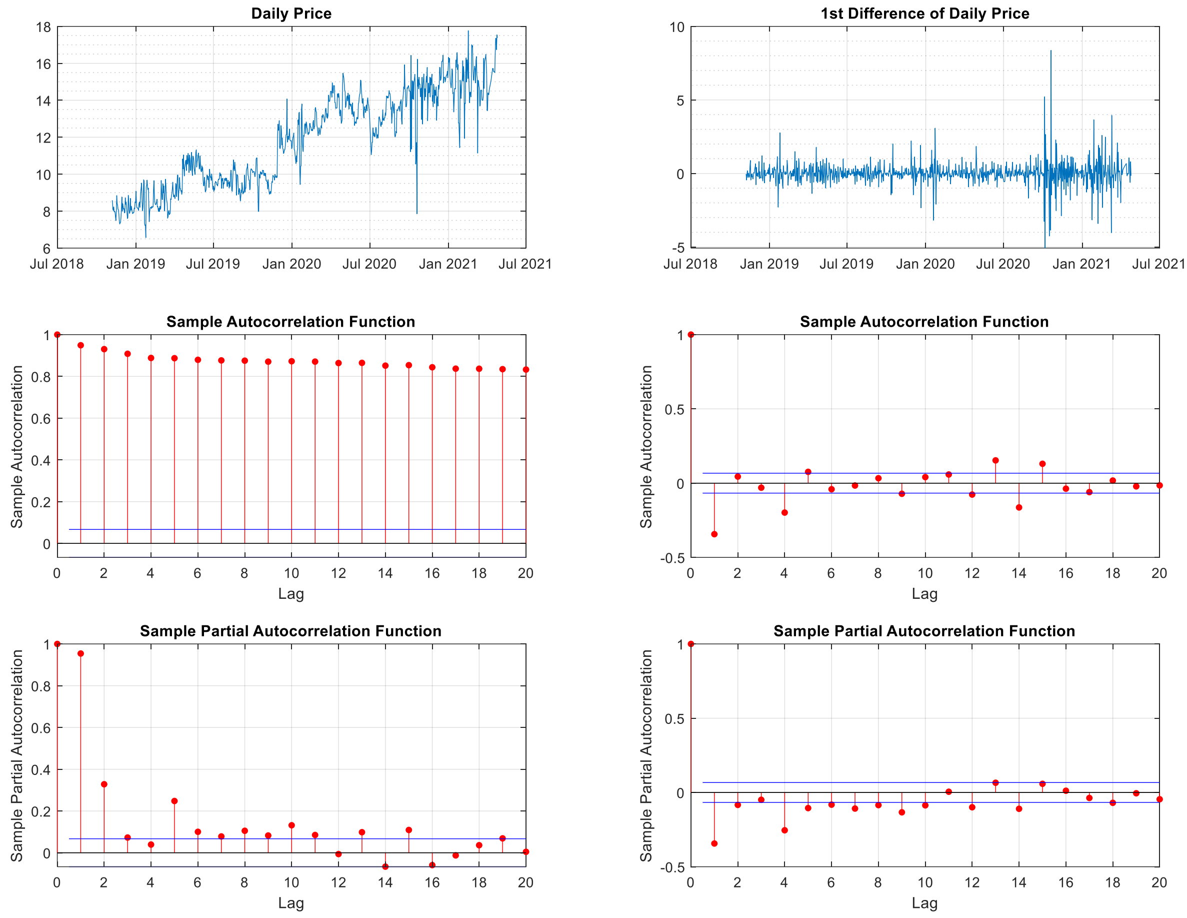
4.2.1 Stationary inspection

The stationarity analysis of the time series belonging to the fruit category was carried out. Graphics are presented in Figure 5. On the left side of the figure, the original series is depicted, while on the right side, a first-order differenced series is presented.

It is understood from the graph in the upper left part of the figure that there is an upward trend. The series is not stationary. In the ACF graph just below, the autocorrelation does not reach zero. Therefore, the MA value (q) cannot be determined. In the PACF graph just below, it turns out that the correlation is zero after the second value.

On the right side of the figure, there are cases where difference is taken at the first level. It is possible to say visually that the series becomes stationary when the difference is taken at the first level. It can be stated that the q value is determined as 5 from the ACF graph and the p value is 2 from the PACF graph.

Augmented Dickey Fuller test was used to test whether the time series is stationary. Three models (no constant and no trend, constant included and both constant and trend included) are run separately. The results are in Table 3. It turns out that few categories are stationary in the original version of the series. However, when the first difference is taken, it is understood that the time series of each category is stationary. Therefore, the analysis was continued by taking the first differences.



(a) (b)

Fig. 5. ACF and PACF Graphs for Fruit category (a) level (b) first difference

Table 3
ADF unit root results for each category

	Level						1st Difference		
	ADF Stats			p values			ADF Stats*		
Categories	N	C	CT	N	C	CT	N	C	CT
Carbonated Beverage	-0.30	-2.81	-3.07	0.54	0.06	0.11	-32.12	-32.10	-32.09
Still Drink	0.19	-2.84	-7.83	0.72	0.05	0	-40.54	-40.53	-40.52
Mineral Water	0.07	-2.85	-8.50	0.68	0.05	0	-35.31	-35.29	-35.27
Water	-0.13	-2.86	-4.55	0.60	0.05	0	-29.48	-29.47	-29.45
Milk	0.42	-3.14	-5.18	0.80	0.02	0	-32.95	-32.95	-32.93
Instant Meal	0.62	-2.38	-5.36	0.85	0.15	0	-36.66	-36.68	-36.67
Frozen Food	-0.35	-4.76	-5.16	0.52	0.00	0	-32.30	-32.28	-32.26
Vegetables	-0.16	-4.33	-6.93	0.59	0.00	0	-34.68	-34.66	-34.64
Fruits	-0.64	-4.50	-12.16	0.42	0.00	0	-41.61	-41.59	-41.57
Yogurt	0.42	-2.20	-5.81	0.80	0.21	0	-37.38	-37.38	-37.37
Egg	-0.05	-3.75	-5.41	0.63	0.00	0	-34.45	-34.44	-34.43
Breakfast	0.17	-4.02	-7.80	0.71	0.00	0	-39.23	-39.21	-39.19
Olive	0.58	-1.93	-5.60	0.84	0.33	0.00	-30.51	-30.52	-30.51
Cheese	0.97	-1.19	-5.89	0.91	0.66	0.00	-37.42	-37.47	-37.46
Sugar	0.36	-2.24	-3.11	0.78	0.19	0.11	-33.85	-33.84	-33.82

Table 3
Continued

	Level						1st Difference		
	ADF Stats			p values			ADF Stats*		
Salt and Spices	-0.52	-1.96	-2.46	0.46	0.32	0.37	-27.28	-27.26	-27.23
Bakery and Patisserie	0.34	-2.32	-7.61	0.78	0.17	0.00	-42.42	-42.42	-42.40
Liquid Oil	0.15	-2.18	-5.31	0.71	0.22	0.00	-36.42	-36.40	-36.40
White Meat	-0.95	-6.81	-11.16	0.30	0.00	0.00	-39.14	-39.12	-39.10
Meat Deli	0.35	-1.88	-6.44	0.78	0.35	0.00	-34.17	-34.17	-34.16
Fish Seafood	0.14	-2.61	-6.24	0.70	0.09	0.00	-40.51	-40.51	-40.48
Oral Care	-0.56	-5.51	-5.51	0.45	0.00	0.00	-32.76	-32.75	-32.73
Washing up	-0.54	-7.07	-10.42	0.45	0.00	0.00	-35.27	-35.25	-35.23
Laundry Utensils	-0.81	-3.19	-3.38	0.35	0.02	0.06	-31.23	-31.22	-31.21
Washing	-0.12	-5.62	-10.68	0.61	0.00	0.00	-30.40	-30.38	-30.37
Shaving Supplies	-0.42	-4.54	-7.92	0.50	0.00	0.00	-32.81	-32.79	-32.77

* for each category and for each model $p < 0.01$
N: No Trend and Constant
C: Include constant
CT: Include constant and trend

4.2.2 White noise test

In order to assess the presence of white noise in the transformed time series, an application of the Ljung-Box Statistics is performed. The resulting p-values, are found to be below the predefined significance level of 0.05. This statistical comparison implies that the null hypothesis, positing the absence of serial correlation (white noise), is to be rejected. Consequently, the analysis indicates that the transformed time series does not exhibit characteristics of white noise. This, in turn, signifies the presence of underlying patterns or dependencies within the data, suggesting the feasibility of constructing a meaningful model for prediction and further analysis. [71].

4.2.3 Determining optimal number of p and q

Although it is possible to determine parameters with ACF and PACF graphics, it is preferred to choose the parameters with the lowest BIC value instead of specifying parameters with the help of graphics each time, since too many ARIMA models need to be run in the study. ARIMA models were run for fruit category with parameters p and q ranging from 1 to 5, and the resulting Bayesian Information Criterion (BIC) values have been reported in Table 4. The minimum BIC value is obtained when the AR degree (p) is 2 and MA degree (q) is 5.

Table 4
BIC values for different components of p and q (Fruit Category)

		Q				
		1	2	3	4	5*
P	1	1851.551	1839.953	1796.074	1769.527	1773.025
	2*	1802.404	1796.287	1782.131	1772.333	1766.276
	3	1791.459	1869.110	1780.514	1768.926	1775.267
	4	1861.134	1858.864	1776.348	1771.361	1776.103
	5	1773.710	1775.101	1778.976	1782.460	1779.865

* The best BIC value is obtained with p = 2 and q = 5

The details of the ARIMA (2,1,5) model for the first forecasted day of fruit category is presented in Table 5. All of the coefficients are statistically significant at 0.10 level except for AR(2) coefficient.

Table 5
ARIMA (2,1,5) Model for the fruit category for the first out-of-sample day forecast

	Value	Standard Error	T Statistics	P value
Constant	0.0163	0.0084	1.9361	0.0529
AR{1}	-0.9006	0.2060	-4.372	0.0000
AR{2}	0.0003	0.1659	0.0020	0.9984
MA{1}	0.4340	0.2065	2.1018	0.0356
MA{2}	-0.4552	0.0768	-5.9234	0.0000
MA{3}	-0.1538	0.0832	-1.8485	0.0645
MA{4}	-0.3514	0.0446	-7.8828	0.0000
MA{5}	-0.1465	0.0719	-2.0379	0.0416
Variance	0.4470	0.0117	38.2380	0.0000

Effective Sample Size : 851
Number of estimated parameters : 9
LogLikelihood : -864.864
AIC : 1747.73
BIC : 1790.45

Equation of the Fruit category for the first day of testing period;

$$y'_t = 0.0163 - 0.9006y'_{t-1} + 0.0003277y'_{t-2} + 0.4340\varepsilon_{t-1} - 0.4552\varepsilon_{t-2} - 0.1538\varepsilon_{t-3} - 0.3514\varepsilon_{t-4} - 0.1465\varepsilon_{t-5}$$

Equation of the Fruit category for the second day of testing period;

$$y'_t = 0.0166 - 0.8976y'_{t-1} + 0.0049y'_{t-2} + 0.4287\varepsilon_{t-1} - 0.4613\varepsilon_{t-2} - 0.1495\varepsilon_{t-3} - 0.3451\varepsilon_{t-4} - 0.1458\varepsilon_{t-5}$$

Equation of the Fruit category for the last day of testing period;

$$y'_t = 0.0166 - 0.9573y'_{t-1} - 0.0470y'_{t-2} + 0.4990\varepsilon_{t-1} - 0.4451\varepsilon_{t-2} - 0.2195\varepsilon_{t-3} - 0.3729\varepsilon_{t-4} - 0.1440\varepsilon_{t-5}$$

The values of the parameters of the fruit category during the test set are shown graphically in Figure 6. According to the figure, although the values of the parameters took different values during the test process, they remained relatively stable.

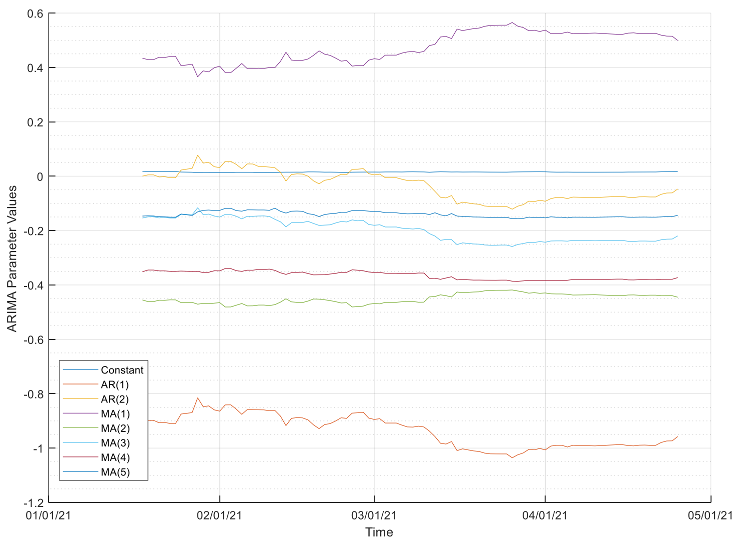


Fig. 6. ARIMA parameters during forecasting interval

Actual values and forecasting values of the model are shown in Figure 7. The performance of the model in the last ninety days is important. because this period is defined as the out-of-sample period.

It is concluded from the graph that the actual values and the estimated values are close to each other. The difference between the actual price and the estimated price fluctuates within the range of ± 2.

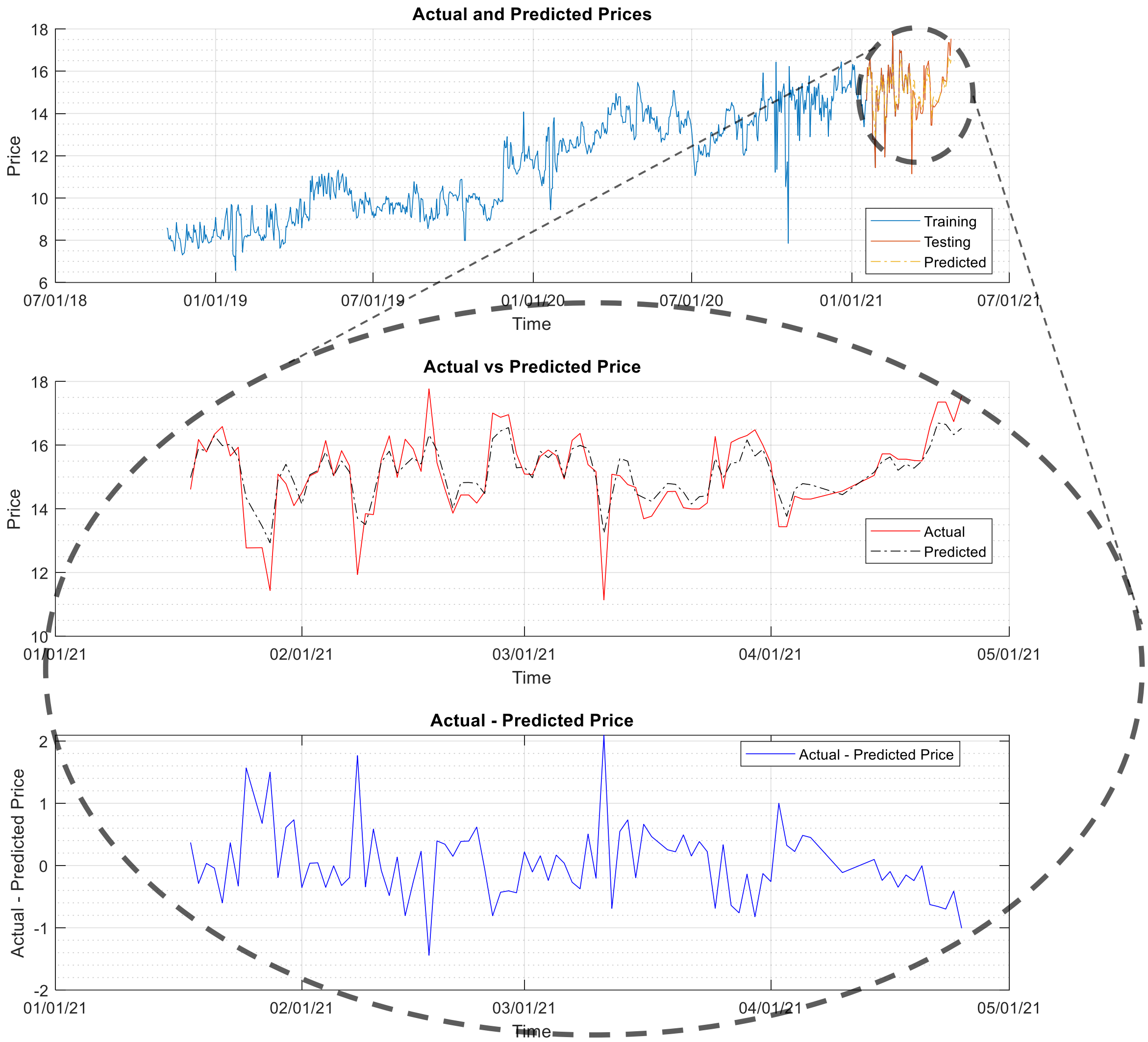


Fig. 7. Actual and forecasted price of the fruit category – Model 1

When the forecasted values of Model 1 and the actual values are examined visually, it is seen that the out-of-sample performance is successful. However, besides visual inspection, numerical measures of the metrics in the performance measurement pool listed in the table are also needed for comparison. In addition to the optimal p and q values for each category, performance measures are also included in Table 6. In Model 1, 26*90 = 2340 ARIMA analysis are run.

Table 6
Performance metrics of ARIMA models

ARIMA models	p	q	RSquared	Adj R-2	Hit Ratio	DS	MAE	MSE	RMSE	MARE	MSRE	RMSRE	MAPE	MSPE	RMSPE	UTheil
Carbonated Beverage	1	1	0.44	0.43	0.49	0.46	0.11	0.02	0.16	0.01	0.000	0.02	1.34	0.04	0.20	0.01
Still Drink	1	1	0.52	0.51	0.48	0.46	0.11	0.02	0.14	0.01	0.000	0.02	1.46	0.04	0.19	0.01
Mineral Water	3	3	0.49	0.49	0.47	0.44	0.16	0.05	0.23	0.02	0.001	0.03	1.98	0.08	0.28	0.01
Water	1	1	0.51	0.5	0.49	0.46	0.08	0.01	0.1	0.02	0.001	0.02	1.70	0.06	0.24	0.01
Milk	1	1	0.50	0.49	0.43	0.38	0.14	0.04	0.21	0.01	0.001	0.02	1.45	0.05	0.22	0.01
Instant Meal	2	3	0.57	0.56	0.47	0.40	0.12	0.02	0.16	0.02	0.001	0.02	1.68	0.05	0.22	0.01
Frozen Food	2	1	0.56	0.56	0.43	0.39	0.26	0.14	0.37	0.01	0.001	0.02	1.49	0.05	0.22	0.01
Vegetables	1	4	0.59	0.58	0.47	0.44	0.28	0.22	0.47	0.03	0.003	0.05	2.84	0.28	0.53	0.02
Fruits	2	5	0.28	0.27	0.45	0.42	0.80	1.15	1.07	0.05	0.006	0.08	5.48	0.61	0.78	0.04
Yogurt	2	1	0.70	0.69	0.45	0.43	0.16	0.04	0.21	0.02	0.000	0.02	1.60	0.04	0.21	0.01
Egg	2	1	0.75	0.74	0.48	0.31	0.22	0.10	0.31	0.01	0.000	0.02	1.31	0.03	0.18	0.01
Breakfast	1	4	0.14	0.13	0.43	0.39	0.41	0.29	0.54	0.02	0.001	0.02	1.87	0.06	0.25	0.01
Olive	1	4	0.71	0.70	0.47	0.46	0.34	0.40	0.64	0.01	0.001	0.02	1.24	0.06	0.24	0.01
Cheese	1	3	0.44	0.43	0.36	0.31	0.34	0.21	0.46	0.01	0.000	0.01	0.95	0.02	0.13	0.01
Sugar	1	1	0.92	0.92	0.40	0.20	0.1	0.03	0.17	0.01	0.000	0.02	0.93	0.03	0.16	0.01
Salt and Spices	1	1	0.82	0.82	0.40	0.35	0.37	0.37	0.61	0.03	0.003	0.05	3.06	0.25	0.50	0.03
Bakery and Patisserie	2	1	0.42	0.42	0.37	0.34	0.15	0.03	0.19	0.01	0.000	0.02	1.47	0.03	0.19	0.01
Liquid Oil	4	5	0.85	0.84	0.39	0.35	0.46	0.39	0.62	0.01	0.000	0.01	0.84	0.01	0.11	0.01
White Meat	5	4	0.31	0.3	0.35	0.34	1.18	2.28	1.51	0.05	0.004	0.06	4.84	0.40	0.63	0.03
Meat Deli	5	5	0.86	0.86	0.45	0.40	0.69	0.93	0.96	0.01	0.000	0.02	1.32	0.04	0.19	0.01
Fish Seafood	2	1	0.31	0.3	0.33	0.29	0.84	1.42	1.19	0.02	0.001	0.02	1.63	0.06	0.24	0.01
Oral Care	2	1	0.64	0.64	0.40	0.36	1.90	8.15	2.86	0.05	0.007	0.08	4.90	0.67	0.82	0.04
Washing up	2	1	0.26	0.26	0.38	0.35	1.42	3.33	1.82	0.05	0.004	0.06	4.75	0.38	0.62	0.03
Laundry Utensils	4	4	0.64	0.64	0.48	0.24	0.84	1.70	1.30	0.01	0.000	0.02	1.34	0.04	0.21	0.01
Washing	4	1	0.56	0.55	0.47	0.44	0.55	0.71	0.84	0.02	0.001	0.04	2.23	0.12	0.35	0.02
Shaving Supplies	2	1	0.52	0.51	0.46	0.42	1.73	9.01	3.00	0.03	0.004	0.06	3.49	0.4	0.63	0.03

4.3 Model 2 forecasting results

4.3.1 Forecasting with EMD

In Model 2, the time series is first separated into its components by the EMD algorithm. For the fruit category, 6 IMF and one residual were obtained from the original time series data. Components are presented graphically in Figure 8.

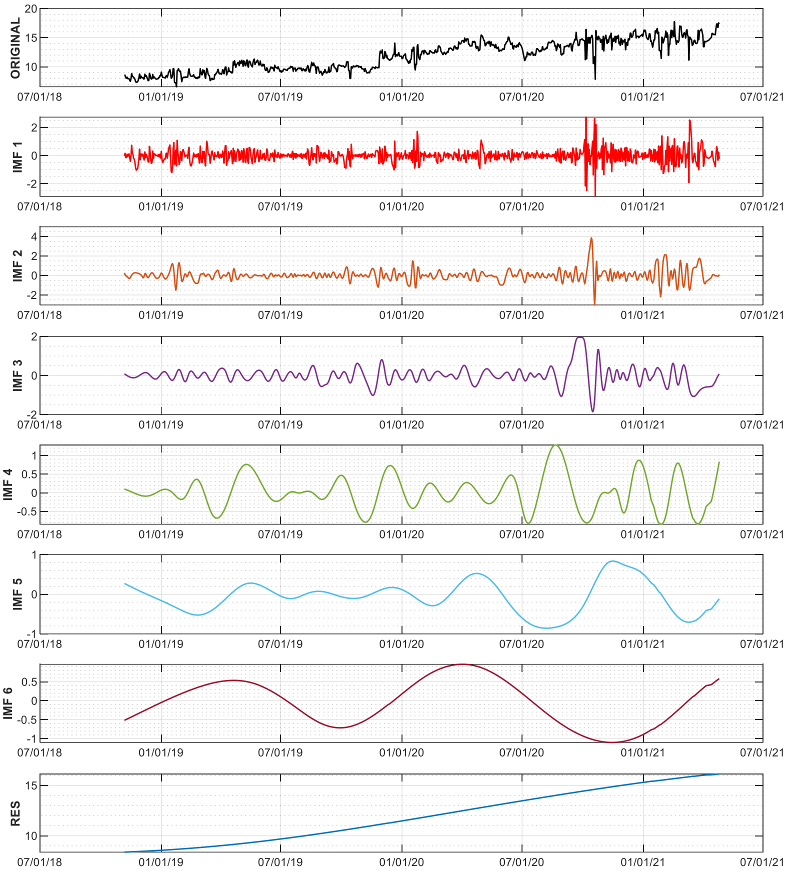


Fig. 8. Signal decomposition with EMD for fruit category

The frequency profiles of the components are presented in Table 7 with the help of descriptive statistics. Whether each component is stationary or not was carried out with the help of the ADF test. The results are in the table. According to the test results, the first three IMFs were found to be stationary in residual level. The remaining components are not stationary at the level. Differentiation was applied to make them stationary. However, each component could become stationary as a result of different number of difference processes. IMF4 and IMF5 became stationary after three times of difference, and IMF6 after four times of difference. As in Model 1, it was tested whether the series were stationary with three different models and the difference process was applied until they were stationary in any model. Thus, the d parameter in the ARIMA (p, d, q) model is optimized for each component, separately. The other parameters in the model, p and q, were determined by searching the smallest BIC value, just like in Model 1. The table lists the parameters for the fruit category. It turns out that different values are determined optimally for each component.

Table 7
Frequency profile of each IMF and residual decomposed by EMD (Fruit category)

		IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	Residual
Descriptive Statistics	Minimum	-2.9014	-3.0276	-1.8545	-0.8458	-0.8563	-1.0998	8.3800
	Maximum	2.7373	3.8634	1.9562	1.2835	0.8379	0.9599	16.1405
	Mean	-0.0077	0.0312	-0.0089	0.0101	-0.0705	-0.0588	11.9186
	%25	-0.2412	-0.1992	-0.2135	-0.2360	-0.3378	-0.5636	9.5653
	Median	-0.0039	-0.0076	-0.0131	-0.0034	-0.0505	-0.0316	11.7572
	%75	0.2018	0.2131	0.1887	0.2608	0.1641	0.4470	14.1714
	Standard Deviation	0.4815	0.5761	0.4583	0.4454	0.4007	0.5945	2.4707
	Skewness	0.3742	1.2296	0.7811	0.3217	0.0351	-0.0735	0.1444
	Kurtosis	9.8451	11.5152	8.2124	3.0809	2.6474	1.9034	1.6511
	JB Stat	1762.2603	2919.7088	1100.4901	15.6296	4.8029	45.4934	70.7306
	p value	0.0010	0.0010	0.0010	0.0022	0.0842	0.0010	0.0010
ADF Test Results	N^0	-34.3463	-8.4124	-3.2795	-1.2806	0.9812	-1.0305	137.7787
	p^0	0.0010	0.0010	0.0011	0.1853	0.9140	0.2743	0.9990
	C^0	-34.3756	-8.4187	-3.2759	-1.2697	0.6551	-0.8080	22.7228
	P^0	0.0010	0.0010	0.0165	0.6199	0.9909	0.8156	0.9990
	CT^0	-34.4482	-8.4657	-3.2753	-1.2643	0.7619	-1.7745	-38.8540
	p^0	0.0010	0.0010	0.0711	0.8948	0.9990	0.7040	0.0010
	N^i				-4.9786	-3.4114	-17.4355	
	P^i				0.0010	0.0010	0.0010	
	C^i				-4.9757	-3.4120	-17.4286	
	P^i				0.0010	0.0110	0.0010	
	CT^i				-4.9729	-3.4182	-17.4381	
	P^i				0.0010	0.0497	0.0010	
	ARIMA p	3	3	4	4	3	1	1
	ARIMA d	0	0	0	3	3	4	0
	ARIMA q	4	5	5	3	1	5	1

^{0}: Level
^{i}: Difference
N: No Trend and Constant
C: Include constant
CT : Include constant and trend

Out of sample estimation results for each component are presented in Figure 9. Although there is a difference between the forecasted values of the model and the actual values in the early days, this difference decreases towards the end of the test set.

It is calculated for the metrics in the performance metric pool for each component and presented in Table 8. Although the performance is relatively low in statistical performance measures, the DS and hit ratio values, which are the estimation measure of the direction of the movement, are at a very good level.

After making the forecasts for each component, it was estimated what the next day's average price in the category would be by taking the sum of these estimates. For the fruit category, the estimated price with the actual price in the test set is plotted in Figure 10.

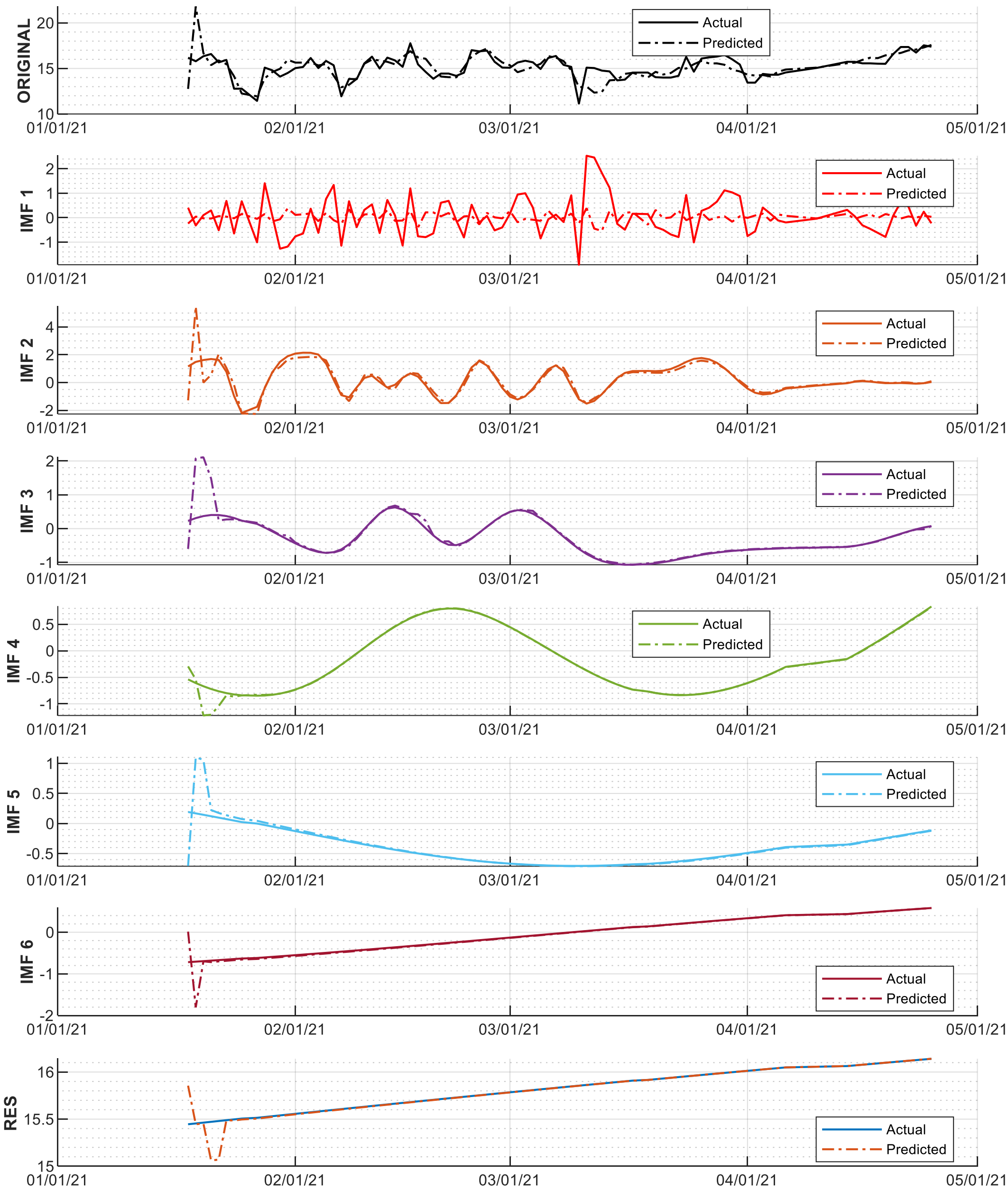


Fig. 9. Out-of-sample forecasting results for each IMF and residual for the fruit category

Table 8
Out-of-sample forecasting results for each component of fruit category

	MAE	MSE	RMSE	MARE	MSRE	RMSRE	MAPE
IMF1	0.63	0.66	0.81	1.09	1.66	1.29	109.48
IMF2	0.24	0.32	0.57	0.64	2.83	1.68	64.23
IMF3	0.09	0.09	0.30	0.52	5.32	2.31	52.29
IMF4	0.02	0.01	0.09	0.05	0.03	0.16	4.84
IMF5	0.04	0.03	0.17	1.28	85.65	9.25	127.51
IMF6	0.03	0.02	0.14	0.13	0.38	0.61	13.11
RES	0.02	0.01	0.07	0.00	0.00	0.00	0.10

Table 8
Continued

	MSPE	RMSPE	UTheil	R^2	Adj R^2	Hit Ratio	DS
IMF1	165.55	12.867	0.83	-0.04	-0.05	0.62	0.62
IMF2	283.00	16.823	0.26	0.69	0.68	0.91	0.91
IMF3	531.65	23.057	0.25	0.61	0.61	0.93	0.93
IMF4	2.65	1.627	0.07	0.98	0.98	0.93	0.93
IMF5	8564.79	92.546	0.17	0.57	0.56	0.99	0.99
IMF6	37.79	6.147	0.17	0.87	0.87	0.98	0.98
RES	0.00	0.048	0.00	0.86	0.86	0.97	0.97

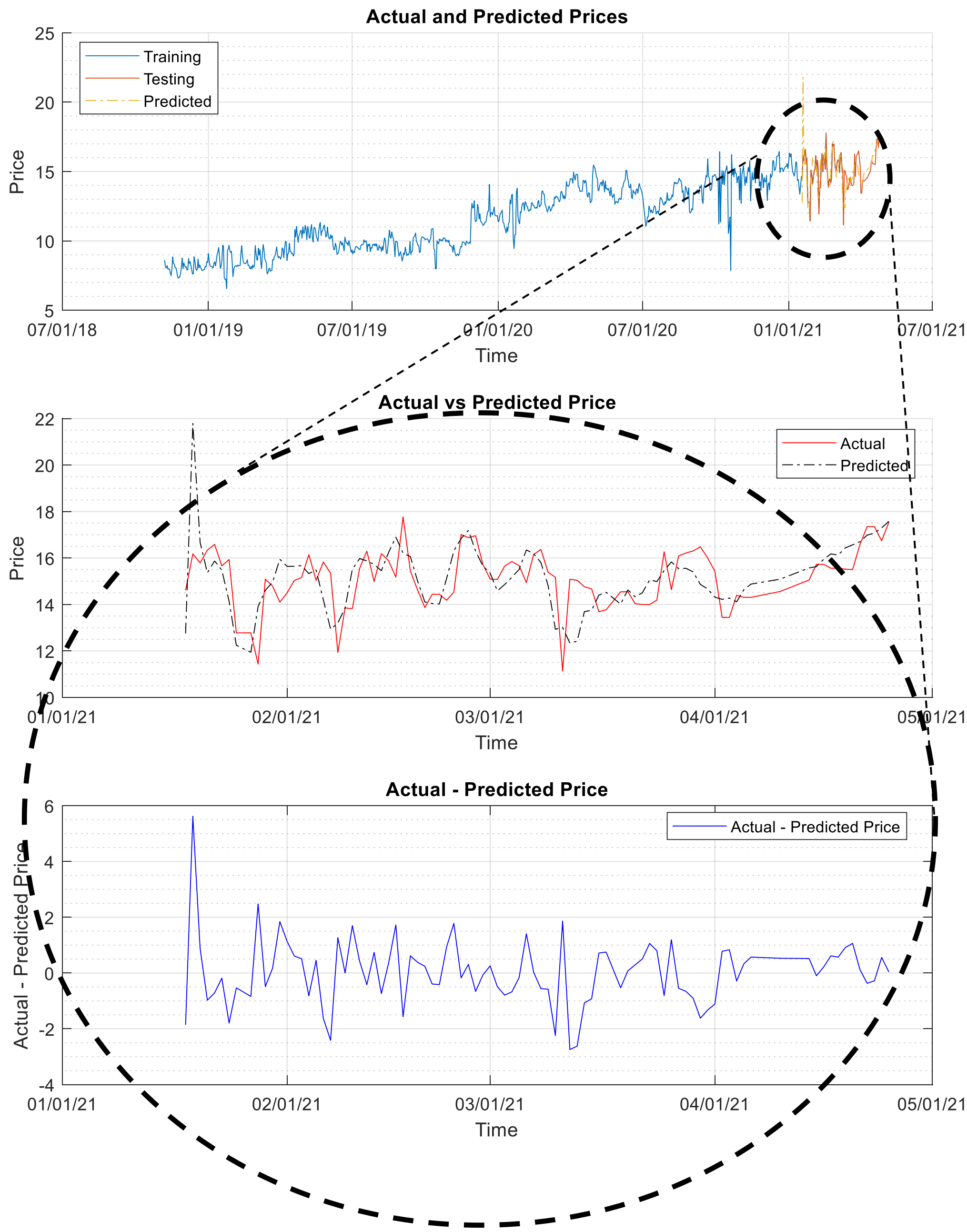


Fig. 10. Actual and forecasted price of the fruit category – Model 2

Each metric in the pool of performance appraisal metrics was calculated with the values estimated by model 2 and the results are presented in Table 9. There are 159 IMFs plus there are 26 residuals so there are total 185 subcomponents. In EMD-ARIMA models there are 185*90*26=432900 ARIMA models are constructed.

Table 9
Out-of-sample forecasting results for each category

Categories	Number of IMFs	R^2	Adj R^2	Hit Ratio	DS	MAE	MSE	RMSE
Carbonated Beverage	7	0.70	0.7	0.51	0.46	0.09	0.01	0.11
Still Drink	6	0.30	0.29	0.61	0.58	0.09	0.03	0.16
Mineral Water	6	0.63	0.63	0.58	0.53	0.14	0.04	0.20
Water	7	0.69	0.69	0.52	0.48	0.06	0.01	0.08
Milk	6	0.43	0.42	0.69	0.65	0.12	0.05	0.22
Instant Meal	6	0.79	0.79	0.70	0.62	0.08	0.01	0.11
Frozen Food	6	0.37	0.37	0.60	0.55	0.24	0.20	0.45
Vegetables	6	0.85	0.85	0.57	0.54	0.20	0.08	0.28
Fruits	6	0.26	0.25	0.60	0.57	0.73	1.18	1.09
Yogurt	6	0.51	0.50	0.61	0.60	0.17	0.07	0.27
Egg	6	0.83	0.83	0.57	0.44	0.17	0.06	0.25
Breakfast	5	0.65	0.65	0.69	0.65	0.27	0.12	0.34
Olive	5	0.85	0.85	0.61	0.58	0.29	0.21	0.45
Cheese	5	0.22	0.21	0.57	0.55	0.28	0.29	0.54
Sugar	5	0.93	0.93	0.55	0.29	0.10	0.03	0.16
Salt and Spices	7	0.73	0.72	0.56	0.53	0.42	0.57	0.75
Bakery and Patisserie	6	0.74	0.74	0.60	0.57	0.10	0.02	0.12
Liquid Oil	6	0.88	0.88	0.52	0.47	0.38	0.30	0.54
White Meat	8	0.57	0.56	0.62	0.60	0.90	1.43	1.20
Meat Deli	6	0.92	0.92	0.65	0.61	0.49	0.49	0.70
Fish Seafood	6	0.58	0.57	0.52	0.44	0.73	0.87	0.93
Oral Care	7	-0.21	-0.22	0.65	0.61	2.16	27.76	5.27
Washing up	6	0.45	0.45	0.63	0.60	1.16	2.48	1.57
Laundry Utensils	6	0.76	0.76	0.61	0.40	0.74	1.12	1.06
Washing	7	0.71	0.71	0.57	0.54	0.45	0.46	0.68
Shaving Supplies	6	0.69	0.69	0.66	0.65	1.51	5.71	2.39

Categories	Number of IMFs	MARE	MSRE	RMSRE	MAPE	MSPE	RMSPE	UTheil
Carbonated Beverage	7	0.01	0.000	0.01	1.09	0.02	0.14	0.01
Still Drink	6	0.01	0.001	0.02	1.20	0.05	0.23	0.01
Mineral Water	6	0.02	0.001	0.02	1.68	0.06	0.24	0.01
Water	7	0.01	0.000	0.02	1.42	0.03	0.19	0.01
Milk	6	0.01	0.001	0.02	1.30	0.06	0.24	0.01
Instant Meal	6	0.01	0.000	0.02	1.12	0.02	0.15	0.01
Frozen Food	6	0.01	0.001	0.03	1.36	0.06	0.25	0.01
Vegetables	6	0.02	0.001	0.03	1.97	0.09	0.30	0.01
Fruits	6	0.05	0.005	0.07	4.88	0.52	0.72	0.04
Yogurt	6	0.02	0.001	0.03	1.66	0.07	0.26	0.01
Egg	6	0.01	0.000	0.02	1.00	0.02	0.15	0.01

Table 9
Continued

Categories	Number of IMFs	MARE	MSRE	RMSRE	MAPE	MSPE	RMSPE	UTheil
Breakfast	5	0.01	0.000	0.02	1.24	0.02	0.16	0.01
Olive	5	0.01	0.000	0.02	1.10	0.03	0.17	0.01
Cheese	5	0.01	0.000	0.02	0.81	0.02	0.16	0.01
Sugar	5	0.01	0.000	0.02	0.98	0.02	0.16	0.01
Salt and Spices	7	0.04	0.004	0.06	3.50	0.36	0.60	0.03
Bakery and Patisserie	6	0.01	0.000	0.01	0.98	0.02	0.12	0.01
Liquid Oil	6	0.01	0.000	0.01	0.69	0.01	0.10	0.01
White Meat	8	0.04	0.003	0.05	3.67	0.25	0.50	0.02
Meat Deli	6	0.01	0.000	0.01	0.95	0.02	0.14	0.01
Fish Seafood	6	0.01	0.000	0.02	1.41	0.03	0.18	0.01
Oral Care	7	0.06	0.031	0.18	6.32	3.09	1.76	0.07
Washing up	6	0.04	0.003	0.05	3.85	0.26	0.51	0.03
Laundry Utensils	6	0.01	0.000	0.02	1.17	0.03	0.17	0.01
Washing	7	0.02	0.001	0.03	1.84	0.08	0.28	0.01
Shaving Supplies	6	0.03	0.003	0.05	3.13	0.27	0.52	0.02
Total	159							

The distribution of optimally determined parameters for ARIMA models run for each category using the EMD-ARIMA model is presented in Table 10. According to the data in the table, among the examined models, the AR parameter has a maximum value of 5, while the MA parameter also has a maximum value of 5. The differencing parameter, on the other hand, takes a maximum value of 4, but in more than half of the optimal models (58.38%), it takes the value of 0 (indicating stationarity at the level).

Table 10
Frequency table of optimal parameters

Parameters								
p			q			d		
Value	Count	Percent	Value	Count	Percent	Value	Count	Percent
1	45	24.32	1	69	37.30	0	108	58.38
2	54	29.19	2	34	18.38	1	9	4.86
3	23	12.43	3	38	20.54	2	2	1.08
4	47	25.41	4	25	13.51	3	34	18.38
5	16	8.65	5	19	10.27	4	32	17.30

When the results of Model 1 and Model 2 are compared, it can be stated that Model 2 produces better results. however, to test whether this difference is statistically significant, a one-sided Wilcoxon signed rank test was performed. Since there are two different performance measures, two different hypotheses have been established. The hypotheses and their results are presented in Figure 11. The figure also presents histograms of the logarithm of performance measures (negative performance metrics are ignored). As a result, it has been proven that Model 2 performs statistically higher than Model 1 in both higher-is-better and smaller-is-better metrics.

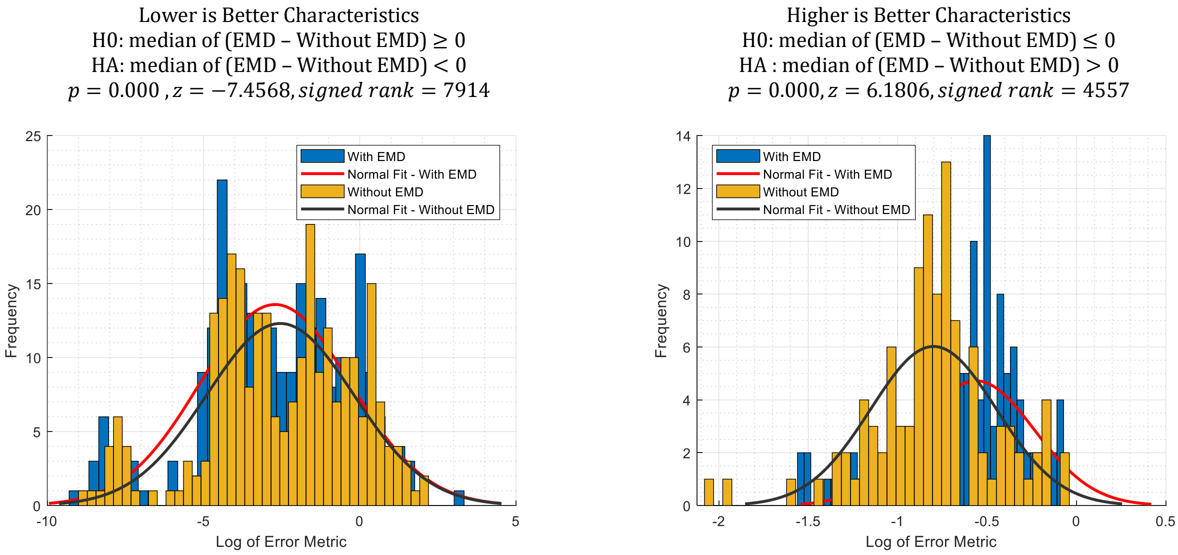


Fig. 11. Two sample Wilcoxon sign rank test results

4.3.2 Relation with standard deviation and performance metrics

In this section, we undertake an investigation into the relationship between the forecasted values generated by Model 2 and the standard deviation values associated with various product categories. Specifically, we employ linear regression analysis to explore the connections between each predictive performance metric and the standard deviation values, presenting our findings graphically in Figure 12. The horizontal axis of the figure represents the standard deviation values of the categories, while the vertical axis conveys the metric values. The figure also incorporates the linear regression equations, with the standard deviation as the independent variable and forecasting performance as the dependent variable.

Our examination focuses on the first four error metrics (the first row of Figure 12), characterized by a “higher-is-better” trait. Within this context, our analysis reveals an inverse relationship between standard deviation and performance metrics. In simpler terms, a reduced standard deviation within a category correlates with superior forecasting performance.

Conversely, the ten graphics featured in the lower section of the figure pertain to metrics where a “smaller-is-better” characteristic prevails within the performance metric pool. Upon close examination, both graphically and through mathematical equations, it becomes evident that a direct linear relationship exists between the standard deviation of the category and the precision of the forecasting process. In essence, our findings indicate that lower category standard deviations are associated with more accurate forecasting, thereby resulting in forecasts that closely align with actual values.

The findings from this analysis yield crucial insights into the intricate relationship between performance metrics, standard deviation, and the integration of Empirical Mode Decomposition (EMD) into Auto Regressive Integrated Moving Average (ARIMA) models. Notably, it was observed that performance metrics characterized by “smaller-is-better” attributes exhibit stronger linear relationships with the standard deviation of product categories, as indicated by higher r^2 values. This underscores the significance of category stability in achieving precise forecasts, where lower standard deviations are linked to more accurate forecasting. Conversely, certain metrics, such as Hit Ratio and DS, showed minimal to negligible linear associations with standard deviation, suggesting that these metrics may be less influenced by category variability.

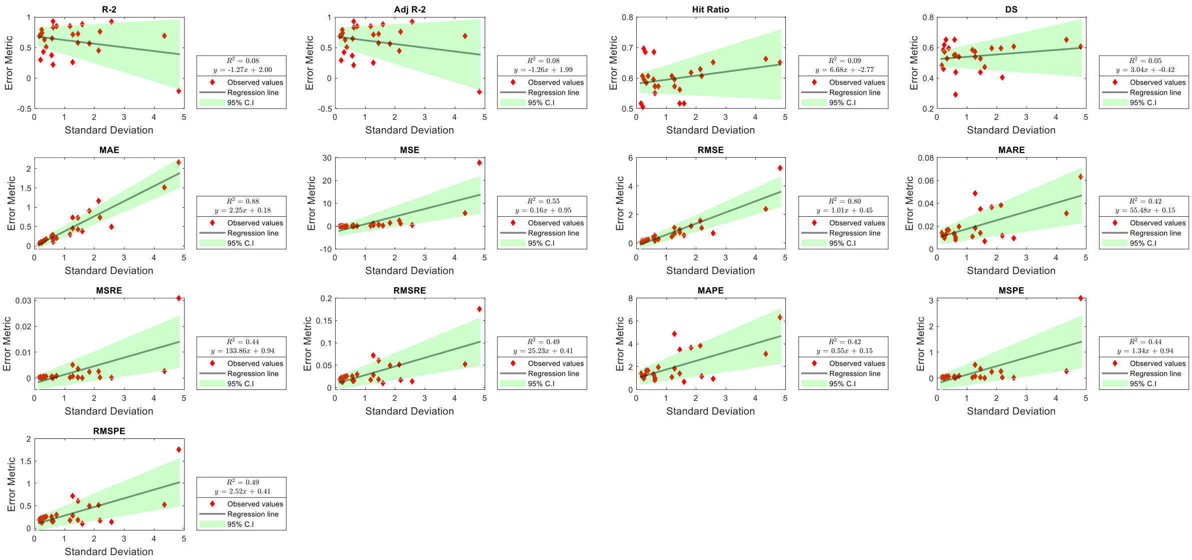


Fig. 12. Error performance metrics and standard deviation

The study also highlights the substantial and statistically significant performance improvement achieved through the incorporation of EMD into the single ARIMA model. This finding aligns with existing literature, reinforcing the utility of EMD in enhancing forecasting accuracy. Additionally, the analysis unveils valuable insights into optimal modeling parameters, with a right-skewed distribution of p-q values suggesting that a limited number of degrees are sufficient for modeling sub-components of time series data. Furthermore, the research pinpoints a maximum difference of four times as the threshold for rendering subcomponents stationary, providing valuable guidance for data transformation processes. The paper's examination of rolling window forecasting models underscores the stability of the constant coefficient across a range of ARIMA models, with other parameters demonstrating minor fluctuations. These findings collectively contribute to a deeper understanding of factors influencing forecasting accuracy and the utility of hybrid models in time series analysis.

4.4 Performance comparison with machine learning algorithm

In this study, addition to the ARIMA model, machine learning algorithms are also utilized to assess the performance of the proposed model. Machine learning algorithms find extensive applications across diverse domains. The effective integration of a machine learning model into varied problem domains necessitates the fine-tuning of its hyper-parameters. The selection of an optimal hyper-parameter configuration significantly influences the performance of the machine learning model. Achieving this optimization often demands a profound understanding of machine learning algorithms and the application of suitable hyper-parameter optimization techniques. While various automatic optimization techniques are available, their efficacy varies, and they exhibit distinct strengths and limitations depending on the specific nature of the problems to which they are applied [73]. For each category and for each subcomponent of the time series, the ML algorithms namely Neural Network (NN), Support Vector Regression (SVR), Regression Tree (RT), Gaussian Process Regression (GPR), and Generalized Additive Model (GAM) is utilized. The parameters of the each method and search range is presented in Table 11.

Table 11
Parameters of the Machine Learning algorithms and search ranges

	Parameter	Search Range	Type
NN	Number of hidden layers	[1 3]	Integer
	Neurons in the hidden layers	[1 300]	Integer
	Activation Functions	[‘relu’,’tanh’,’sigmoid’,’purelinear’]	Categorical
	Standardize inputs	[‘yes’,’no’]	Categorical
	Lambda	[1.2970e-08 129.7017]	Real
	Initializer of Layer Weights	[‘glorot’,’he’]	Categorical
	Initializer of Layer Biases	[‘zeros’,’ones’]	Categorical
SVR	Box Constraint	[1.00e-03 1000]	Real
	Kernel Scale	[1.00e-03 1000]	Real
	Epsilon	[0.0056 564.5173]	Real
	Kernel Function	[‘gaussian’,’linear’,’polynomial’]	Categorical
	Polynomial Order	[2 4]	Integer
	Standardize	[‘true’,’false’]	Categorical
RT	Minimum Leaf Size	[1 385]	Integer
	Maximum Number of Splits	[1 770]	Integer
	Number of Variables to Sample	[1 2]	Integer
GPR	Sigma	[1.e-04 64.0189]	real
	Basis Function	[‘constant’,’none’,’linear’,’pure Quadratic’]	categorical
	Kernel Function	['ardexponential', 'ardmatern32', 'ardmatern52', 'ardrationalquadratic', 'ardsquaredexponential', 'exponential', 'matern32', 'matern52', 'rationalquadratic', 'squaredexponential']	categorical
	Kernel Scale	[1.e-03 1000]	real
	Standardize	[‘yes’,’no’]	categorical
GAM	Initial Learning Rate for Predictors	[1.e-03 1]	real
	Maximum Number of Splits per predictor	[1 30]	Integer
	Number of Trees per predictor	[10 500]	Integer

For each machine learning method and for each sub components of the time series, the optimization process is run and optimal parameters are determined. However, listing all of the optimal parameters will increase the volume of the study, that is why activation functions and hidden layer sizes are reported for the neural networks in Table 12.

Table 12
Optimized activation functions (AF) and hidden layer sizes (HLS) of the Neural Network Model

		Residual	IMF1	IMF2	IMF3	IMF4	...	IMF8
Carbonated Beverage	AF	pure linear	pure linear	pure linear	relu	sigmoid	...	[]
	HLS	[24,2]	[6,157,21]	[161,228]	5	[1,34]		[]
Still Drink	AF	sigmoid	sigmoid	pure linear	pure linear	pure linear		[]
	HLS	4	[20,2]	[1,213]	[3,2]	[1,1]		[]
Mineral Water	AF	relu	sigmoid	pure linear	pure linear	relu		[]
	HLS	17	[296,17,8]	[14,16]	[45,1,6]	16		[]
Water	AF	pure linear	pure linear	pure linear	pure linear	pure linear		[]
	HLS	[73,1,1]	[46,215]	32	[11,10]	[280,2]		[]
Milk	AF	pure linear	sigmoid	pure linear	pure linear	sigmoid		[]
	HLS	[281,1]	26	[18,62]	[1,194,1]	[50,276]		[]
Instant Meal	AF	pure linear	pure linear	relu	pure linear	pure linear		[]
	HLS	[1,180]	[2,241]	59	1	[1,292]		[]
Frozen Food	AF	relu	tanh	pure linear	pure linear	pure linear		[]
	HLS	[29,24]	2	130	[77,32]	[2,11,129]		[]

Table 12
Continued

		Residual	IMF1	IMF2	IMF3	IMF4	...	IMF8
Vegetables	AF	tanh	relu	sigmoid	pure linear	pure linear		[]
	HLS	1	[105,1,66]	22	[13,2,11]	[33,9,42]		[]
Fruits	AF	pure linear	relu	sigmoid	relu	pure linear		[]
	HLS	2	6	[1,278]	264	[16,39,29]		[]
Yogurt	AF	pure linear	tanh	pure linear	pure linear	pure linear		[]
	HLS	[13,2]	[3,1,7]	[212,3,7]	[3,27,2]	10		[]
Egg	AF	pure linear	tanh	pure linear	sigmoid	tanh		[]
	HLS	[197,10,4]	1	[10,47]	1	1		[]
Breakfast	AF	sigmoid	tanh	tanh	tanh	pure linear		[]
	HLS	[2,1]	2	[98,2]	1	5		[]
Olive	AF	relu	relu	relu	pure linear	pure linear		[]
	HLS	[1,9]	[156,1]	125	[3,4]	[2,3,14]		[]
Cheese	AF	pure linear	pure linear	pure linear	pure linear	pure linear		[]
	HLS	[1,2]	2	[36,20,22]	[5,101]	14		[]
Sugar	AF	pure linear	sigmoid	pure linear	pure linear	tanh		[]
	HLS	[58,2]	[166,3,78]	[201,86,20]	[38,2]	22		[]
Salt and Spices	AF	relu	sigmoid	pure linear	tanh	relu		[]
	HLS	[21,89,14]	[1,5]	5	[79,61]	1		[]
Bakery and Patisserie	AF	pure linear	pure linear	pure linear	pure linear	sigmoid		[]
	HLS	[2,162]	1	[2,5]	7	[51,17,2]		[]
Liquid Oil	AF	relu	tanh	pure linear	pure linear	pure linear		[]
	HLS	21	9	5	[1,14]	[4,72,1]		[]
White Meat	AF	sigmoid	pure linear	pure linear	sigmoid	pure linear		relu
	HLS	[3,3,21]	[177,4]	195	[1,43]	[4,4,8]		3
Meat Deli	AF	sigmoid	sigmoid	pure linear	pure linear	relu		[]
	HLS	[1,4]	16	[119,2]	[2,81,198]	[7,3]		[]
Fish Seafood	AF	relu	tanh	pure linear	pure linear	tanh		[]
	HLS	16	4	[62,10,168]	[2,46]	[105,1,6]		[]
Oral Care	AF	tanh	tanh	pure linear	pure linear	pure linear		[]
	HLS	1	[78,13,6]	[1,11]	62	19		[]
Washing up	AF	pure linear	pure linear	pure linear	pure linear	pure linear		[]
	HLS	[5,30,1]	[16,4,22]	[2,2]	2	[4,85,1]		[]
Laundry Utensils	AF	relu	sigmoid	sigmoid	pure linear	pure linear		[]
	HLS	[1,38]	[13,6,121]	103	12	[9,23,76]		[]
Washing	AF	pure linear	sigmoid	relu	pure linear	relu		[]
	HLS	[116,20]	[1,31,3]	1	[11,1,5]	[2,16,142]		[]
Shaving Supplies	AF	sigmoid	relu	relu	tanh	pure linear		[]
	HLS	166	[2,298]	1	2	[2,1,15]		[]

The performance of the neural network model is presented in Table 13. The provided table illustrates the performance metrics of various neural network models across different product categories. Each row corresponds to a specific product category, and the columns present various performance indicators for the corresponding neural network model. RSquared values range from 0.2943 to 0.9624, indicating the proportion of variance in the dependent variable explained by the model. Adjusted R-2 values, ranging from 0.2863 to 0.962, consider the number of predictors in the model, providing a more accurate measure of goodness of fit. Hit Ratio values, spanning from 0.5056 to 0.7303, represent the accuracy of the model's predictions, particularly relevant for classification tasks.

Table 13
Performance of Neural Network

	R^2	Adj R^2	Hit Ratio	DS	MAE	MSE	RMSE
Carbonated Beverage	0.67	0.67	0.56	0.53	0.09	0.01	0.12
Still Drink	0.72	0.71	0.57	0.56	0.08	0.01	0.10
Mineral Water	0.70	0.70	0.62	0.55	0.13	0.03	0.18
Water	0.32	0.31	0.53	0.49	0.10	0.02	0.12
Milk	0.79	0.78	0.62	0.60	0.10	0.02	0.13
Instant Meal	0.76	0.76	0.73	0.63	0.09	0.01	0.11
Frozen Food	0.72	0.71	0.62	0.57	0.22	0.09	0.30
Vegetables	0.78	0.78	0.67	0.64	0.22	0.11	0.34
Fruits	0.29	0.29	0.55	0.53	0.75	1.13	1.06
Yogurt	0.79	0.79	0.52	0.51	0.14	0.03	0.17
Egg	0.87	0.87	0.58	0.45	0.17	0.05	0.22
Breakfast	0.49	0.49	0.63	0.60	0.32	0.17	0.41
Olive	0.83	0.83	0.58	0.55	0.30	0.23	0.48
Cheese	0.77	0.76	0.63	0.61	0.24	0.09	0.29
Sugar	0.96	0.96	0.54	0.30	0.09	0.01	0.12
Salt and Spices	0.89	0.89	0.54	0.51	0.36	0.22	0.47
Bakery and Patisserie	0.68	0.68	0.63	0.60	0.11	0.02	0.14
Liquid Oil	0.93	0.92	0.51	0.47	0.34	0.19	0.43
White Meat	0.56	0.56	0.64	0.60	0.95	1.44	1.20
Meat Deli	0.90	0.90	0.63	0.60	0.61	0.63	0.80
Fish Seafood	0.64	0.63	0.54	0.46	0.69	0.75	0.87
Oral Care	0.83	0.83	0.67	0.63	1.40	3.80	1.95
Washing up	0.56	0.55	0.61	0.57	1.11	1.99	1.41
Laundry Utensils	0.78	0.78	0.56	0.36	0.74	1.02	1.01
Washing	0.69	0.69	0.62	0.61	0.44	0.50	0.71
Shaving Supplies	0.69	0.69	0.57	0.56	1.48	5.77	2.40
	MARE	MSRE	RMSRE	MAPE	MSPE	RMSPE	UTheil
Carbonated Beverage	0.01	0.000	0.02	1.12	0.02	0.15	0.01
Still Drink	0.01	0.000	0.01	1.13	0.02	0.14	0.01
Mineral Water	0.02	0.001	0.02	1.61	0.05	0.21	0.01
Water	0.02	0.001	0.03	2.16	0.07	0.27	0.01
Milk	0.01	0.000	0.01	1.05	0.02	0.14	0.01
Instant Meal	0.01	0.000	0.02	1.26	0.03	0.16	0.01
Frozen Food	0.01	0.000	0.02	1.29	0.03	0.17	0.01
Vegetables	0.02	0.001	0.04	2.18	0.12	0.35	0.02
Fruits	0.05	0.006	0.08	5.13	0.57	0.75	0.03
Yogurt	0.01	0.000	0.02	1.37	0.03	0.17	0.01
Egg	0.01	0.000	0.01	0.98	0.02	0.13	0.01
Breakfast	0.01	0.000	0.02	1.46	0.04	0.19	0.01
Olive	0.01	0.000	0.02	1.12	0.03	0.18	0.01
Cheese	0.01	0.000	0.01	0.67	0.01	0.08	0.00
Sugar	0.01	0.000	0.01	0.80	0.01	0.11	0.01
Salt and Spices	0.03	0.002	0.04	3.03	0.16	0.40	0.02
Bakery and Patisserie	0.01	0.000	0.01	1.10	0.02	0.14	0.01
Liquid Oil	0.01	0.000	0.01	0.62	0.01	0.08	0.00
White Meat	0.04	0.002	0.05	3.83	0.23	0.48	0.02

Table 13
Continued

	MARE	MSRE	RMSRE	MAPE	MSPE	RMSPE	UTheil
Meat Deli	0.01	0.000	0.02	1.15	0.02	0.15	0.01
Fish Seafood	0.01	0.000	0.02	1.33	0.03	0.17	0.01
Oral Care	0.04	0.002	0.05	3.51	0.24	0.49	0.02
Washing up	0.04	0.002	0.05	3.70	0.23	0.47	0.02
Laundry Utensils	0.01	0.000	0.02	1.18	0.03	0.16	0.01
Washing	0.02	0.001	0.03	1.78	0.08	0.29	0.01
Shaving Supplies	0.03	0.002	0.05	2.90	0.21	0.46	0.02

The performance of the Support Vector Regression is presented in Table 14. RSquared values range from 0.4229 to 0.9622, indicating the proportion of the dependent variable's variance explained by the models. Higher values suggest better fitting models. Adjusted R-2, which accounts for the number of predictors, follows a similar pattern, showing the models perform relatively well in capturing variance. Hit Ratio and DS values vary across categories, indicating the predictive accuracy and discrimination ability of the models for different product types. Lower values for MAE, MSE, RMSE, MAPE, MSPE, and RMSPE generally indicate better predictive performance. UTheil values measure the forecasting accuracy relative to a benchmark model. Smaller UTheil coefficients suggest better forecasting performance. Products like "Sugar" and "Liquid Oil" exhibit very high RSquared values, suggesting that the models for these categories are effective in explaining variance. "Fruits" have relatively lower RSquared and higher error metrics, indicating potential challenges in predicting sales for this category. The models generally perform well across various product categories, but there are variations, and understanding the specific context of each category is crucial for accurate interpretation.

Table 14
Performance of Support Vector Regression

	R^2	Adj R^2	Hit Ratio	DS	MAE	MSE	RMSE
Carbonated Beverage	0.67	0.67	0.55	0.52	0.09	0.01	0.12
Still Drink	0.73	0.72	0.61	0.60	0.08	0.01	0.10
Mineral Water	0.71	0.71	0.60	0.54	0.13	0.03	0.18
Water	0.68	0.68	0.56	0.54	0.06	0.01	0.08
Milk	0.77	0.77	0.60	0.57	0.10	0.02	0.14
Instant Meal	0.76	0.76	0.74	0.64	0.09	0.01	0.12
Frozen Food	0.70	0.70	0.60	0.55	0.23	0.09	0.31
Vegetables	0.74	0.74	0.62	0.57	0.23	0.14	0.37
Fruits	0.42	0.42	0.58	0.56	0.74	0.92	0.96
Yogurt	0.79	0.79	0.52	0.51	0.14	0.03	0.18
Egg	0.87	0.86	0.61	0.46	0.17	0.05	0.23
Breakfast	0.51	0.50	0.66	0.63	0.32	0.17	0.41
Olive	0.82	0.81	0.54	0.51	0.35	0.26	0.51
Cheese	0.76	0.76	0.63	0.61	0.24	0.09	0.30
Sugar	0.96	0.96	0.52	0.28	0.09	0.01	0.12
Salt and Spices	0.89	0.88	0.54	0.51	0.34	0.24	0.49
Bakery and Patisserie	0.68	0.68	0.63	0.60	0.11	0.02	0.14
Liquid Oil	0.92	0.92	0.49	0.45	0.36	0.20	0.44
White Meat	0.57	0.56	0.61	0.56	0.97	1.42	1.19
Meat Deli	0.92	0.92	0.63	0.60	0.54	0.52	0.72
Fish Seafood	0.63	0.63	0.53	0.44	0.68	0.76	0.87
Oral Care	0.82	0.82	0.67	0.63	1.43	4.10	2.02
Washing up	0.57	0.57	0.65	0.62	1.09	1.93	1.39

Table 14
Continued

	R^2	Adj R^2	Hit Ratio	DS	MAE	MSE	RMSE
Laundry Utensils	0.77	0.77	0.56	0.35	0.76	1.08	1.04
Washing	0.70	0.70	0.58	0.57	0.43	0.48	0.69
Shaving Supplies	0.69	0.69	0.58	0.57	1.46	5.75	2.40
	MARE	MSRE	RMSRE	MAPE	MSPE	RMSPE	UTheil
Carbonated Beverage	0.01	0.000	0.02	1.12	0.02	0.15	0.01
Still Drink	0.01	0.000	0.01	1.11	0.02	0.14	0.01
Mineral Water	0.02	0.000	0.02	1.59	0.04	0.21	0.01
Water	0.01	0.000	0.02	1.46	0.04	0.19	0.01
Milk	0.01	0.000	0.01	1.08	0.02	0.14	0.01
Instant Meal	0.01	0.000	0.02	1.29	0.03	0.17	0.01
Frozen Food	0.01	0.000	0.02	1.33	0.03	0.18	0.01
Vegetables	0.02	0.002	0.04	2.35	0.16	0.40	0.02
Fruits	0.05	0.005	0.07	5.03	0.46	0.68	0.03
Yogurt	0.01	0.000	0.02	1.39	0.03	0.17	0.01
Egg	0.01	0.000	0.01	0.99	0.02	0.13	0.01
Breakfast	0.01	0.000	0.02	1.45	0.03	0.19	0.01
Olive	0.01	0.000	0.02	1.28	0.04	0.19	0.01
Cheese	0.01	0.000	0.01	0.68	0.01	0.08	0.00
Sugar	0.01	0.000	0.01	0.80	0.01	0.11	0.01
Salt and Spices	0.03	0.002	0.04	2.85	0.16	0.40	0.02
Bakery and Patisserie	0.01	0.000	0.01	1.10	0.02	0.14	0.01
Liquid Oil	0.01	0.000	0.01	0.65	0.01	0.08	0.00
White Meat	0.04	0.002	0.05	3.88	0.23	0.48	0.02
Meat Deli	0.01	0.000	0.01	1.02	0.02	0.14	0.01
Fish Seafood	0.01	0.000	0.02	1.31	0.03	0.17	0.01
Oral Care	0.04	0.003	0.05	3.60	0.26	0.51	0.03
Washing up	0.04	0.002	0.05	3.63	0.22	0.47	0.02
Laundry Utensils	0.01	0.000	0.02	1.20	0.03	0.16	0.01
Washing	0.02	0.001	0.03	1.73	0.08	0.28	0.01
Shaving Supplies	0.03	0.002	0.05	2.88	0.21	0.46	0.02

Forecasting performance of the Regression Tree is presented in Table 15. The RSquared values for the Regression Tree models range from 0.3305 to 0.9616, showing considerable variability in the explained variance across different product categories. Adjusted R-2 values follow a similar pattern, suggesting the importance of predictors in improving model fit. Hit Ratio and DS values also exhibit variations, indicating the model's ability to predict the outcome accurately and discriminate between different categories. Similar to the Neural Network models, these metrics provide insights into the accuracy and precision of the Regression Tree models. Lower values are generally desirable. UTheil coefficients provide a measure of forecasting accuracy relative to a benchmark model. Smaller values suggest better forecasting performance. "Sugar" and "Liquid Oil" categories continue to show high RSquared values, indicating effective model performance. "Fruits" still exhibit challenges with lower RSquared and higher error metrics compared to other categories.

Table 15
Performance of Regression Tree

	R^2	Adj R^2	Hit Ratio	DS	MAE	MSE	RMSE
Carbonated Beverage	0.64	0.64	0.58	0.54	0.09	0.02	0.12
Still Drink	0.69	0.69	0.62	0.60	0.08	0.01	0.11
Mineral Water	0.70	0.70	0.54	0.49	0.14	0.03	0.18
Water	0.71	0.71	0.54	0.51	0.06	0.01	0.08
Milk	0.71	0.70	0.61	0.60	0.12	0.03	0.16
Instant Meal	0.76	0.76	0.71	0.63	0.09	0.01	0.11
Frozen Food	0.75	0.74	0.57	0.53	0.22	0.08	0.28
Vegetables	0.73	0.72	0.61	0.56	0.27	0.14	0.38
Fruits	0.33	0.32	0.56	0.54	0.76	1.07	1.03
Yogurt	0.72	0.72	0.58	0.57	0.16	0.04	0.20
Egg	0.75	0.74	0.64	0.46	0.25	0.10	0.31
Breakfast	0.41	0.40	0.64	0.60	0.35	0.20	0.45
Olive	0.72	0.71	0.54	0.49	0.41	0.39	0.63
Cheese	0.62	0.62	0.60	0.57	0.30	0.14	0.38
Sugar	0.96	0.96	0.57	0.29	0.09	0.01	0.12
Salt and Spices	0.82	0.82	0.55	0.51	0.50	0.38	0.62
Bakery and Patisserie	0.64	0.63	0.67	0.64	0.11	0.02	0.15
Liquid Oil	0.79	0.78	0.56	0.52	0.54	0.54	0.73
White Meat	0.55	0.55	0.62	0.56	0.98	1.48	1.22
Meat Deli	0.71	0.70	0.58	0.55	1.04	1.91	1.38
Fish Seafood	0.58	0.58	0.52	0.44	0.75	0.87	0.93
Oral Care	0.84	0.84	0.64	0.57	1.39	3.72	1.93
Washing up	0.59	0.58	0.58	0.55	1.04	1.86	1.36
Laundry Utensils	0.75	0.75	0.46	0.26	0.79	1.16	1.08
Washing	0.66	0.66	0.61	0.60	0.47	0.54	0.73
Shaving Supplies	0.62	0.62	0.55	0.53	1.83	7.04	2.65
	MARE	MSRE	RMSRE	MAPE	MSPE	RMSPE	UTheil
Carbonated Beverage	0.01	0.000	0.02	1.16	0.02	0.16	0.01
Still Drink	0.01	0.000	0.01	1.15	0.02	0.15	0.01
Mineral Water	0.02	0.001	0.02	1.68	0.05	0.22	0.01
Water	0.01	0.000	0.02	1.46	0.03	0.18	0.01
Milk	0.01	0.000	0.02	1.20	0.03	0.16	0.01
Instant Meal	0.01	0.000	0.02	1.24	0.03	0.16	0.01
Frozen Food	0.01	0.000	0.02	1.27	0.03	0.17	0.01
Vegetables	0.03	0.002	0.04	2.66	0.16	0.40	0.02
Fruits	0.05	0.005	0.07	5.16	0.53	0.72	0.03
Yogurt	0.02	0.000	0.02	1.52	0.04	0.20	0.01
Egg	0.01	0.000	0.02	1.43	0.03	0.18	0.01
Breakfast	0.02	0.000	0.02	1.59	0.04	0.20	0.01
Olive	0.02	0.001	0.02	1.51	0.05	0.23	0.01
Cheese	0.01	0.000	0.01	0.83	0.01	0.10	0.01
Sugar	0.01	0.000	0.01	0.80	0.01	0.11	0.01
Salt and Spices	0.04	0.003	0.06	4.32	0.30	0.55	0.03
Bakery and Patisserie	0.01	0.000	0.01	1.10	0.02	0.15	0.01
Liquid Oil	0.01	0.000	0.01	0.96	0.02	0.13	0.01
White Meat	0.04	0.002	0.05	3.90	0.23	0.48	0.02
Meat Deli	0.02	0.001	0.03	1.93	0.06	0.25	0.01
Fish Seafood	0.01	0.000	0.02	1.44	0.03	0.18	0.01
Oral Care	0.03	0.002	0.05	3.49	0.23	0.48	0.02

Table 15
Continued

	R^2	Adj R^2	Hit Ratio	DS	MAE	MSE	RMSE
Washing up	0.03	0.002	0.05	3.49	0.22	0.46	0.02
Laundry Utensils	0.01	0.000	0.02	1.25	0.03	0.17	0.01
Washing	0.02	0.001	0.03	1.93	0.09	0.30	0.01
Shaving Supplies	0.04	0.003	0.05	3.61	0.29	0.53	0.03

The performance of the Gaussian Process Regression model is presented in the Table 16. The RSquared values for the Gaussian Process Regression models range from -0.3720 to 0.9685, indicating a broad spectrum of predictive performance across different product categories. Adjusted R-2 values generally follow the same trend. Hit Ratio and DS values vary, showing the ability of the model to predict outcomes accurately and discriminate between different categories. These metrics provide insights into the accuracy and precision of the Gaussian Process Regression models. Lower values are generally desirable. UTheil coefficients provide a measure of forecasting accuracy relative to a benchmark model. Smaller values suggest better forecasting performance. "Sugar" and "Liquid Oil" categories exhibit high RSquared values, indicating effective model performance. "Washing up" category has a negative RSquared value, suggesting a poor fit for this model. The Gaussian Process Regression models show diverse performance across different product categories. The negative RSquared value for "Washing up" suggests that the model may not be suitable for predicting sales in this category.

Table 16
Performance of Gaussian Process Regression

	R^2	Adj R^2	Hit Ratio	DS	MAE	MSE	RMSE
Carbonated Beverage	0.67	0.67	0.57	0.54	0.09	0.01	0.12
Still Drink	0.70	0.70	0.63	0.61	0.08	0.01	0.11
Mineral Water	0.71	0.71	0.76	0.71	0.12	0.03	0.18
Water	0.79	0.78	0.71	0.67	0.05	0.00	0.07
Milk	0.74	0.74	0.78	0.75	0.10	0.02	0.15
Instant Meal	0.75	0.75	0.71	0.61	0.09	0.01	0.12
Frozen Food	0.80	0.79	0.79	0.74	0.19	0.06	0.25
Vegetables	0.76	0.76	0.67	0.63	0.22	0.12	0.35
Fruits	0.46	0.46	0.56	0.54	0.70	0.86	0.93
Yogurt	0.79	0.79	0.52	0.51	0.14	0.03	0.17
Egg	0.87	0.87	0.61	0.46	0.17	0.05	0.22
Breakfast	0.65	0.65	0.84	0.80	0.24	0.12	0.34
Olive	0.81	0.81	0.55	0.52	0.31	0.26	0.51
Cheese	0.86	0.86	0.83	0.79	0.15	0.05	0.23
Sugar	0.97	0.97	0.52	0.30	0.08	0.01	0.11
Salt and Spices	0.92	0.92	0.64	0.60	0.29	0.16	0.40
Bakery and Patisserie	0.79	0.78	0.80	0.78	0.09	0.01	0.11
Liquid Oil	0.93	0.93	0.53	0.48	0.34	0.18	0.43
White Meat	0.36	0.36	0.78	0.73	1.00	2.10	1.45
Meat Deli	0.91	0.90	0.58	0.55	0.59	0.62	0.79
Fish Seafood	0.85	0.85	0.78	0.67	0.36	0.32	0.56
Oral Care	0.85	0.84	0.65	0.61	1.37	3.54	1.88
Washing up	-0.37	-0.39	0.62	0.58	1.90	6.20	2.49
Laundry Utensils	0.76	0.76	0.55	0.38	0.83	1.11	1.05
Washing	0.74	0.73	0.63	0.62	0.44	0.42	0.65
Shaving Supplies	0.70	0.70	0.58	0.57	1.43	5.54	2.35

Table 16
Continued

	MARE	MSRE	RMSRE	MAPE	MSPE	RMSPE	UTheil
Carbonated Beverage	0.01	0.000	0.02	1.14	0.02	0.15	0.01
Still Drink	0.01	0.000	0.01	1.15	0.02	0.15	0.01
Mineral Water	0.02	0.000	0.02	1.50	0.04	0.21	0.01
Water	0.01	0.000	0.02	1.24	0.02	0.15	0.01
Milk	0.01	0.000	0.02	1.01	0.02	0.15	0.01
Instant Meal	0.01	0.000	0.02	1.30	0.03	0.17	0.01
Frozen Food	0.01	0.000	0.01	1.08	0.02	0.15	0.01
Vegetables	0.02	0.001	0.04	2.23	0.14	0.37	0.02
Fruits	0.05	0.004	0.07	4.72	0.44	0.66	0.03
Yogurt	0.01	0.000	0.02	1.37	0.03	0.17	0.01
Egg	0.01	0.000	0.01	0.98	0.02	0.13	0.01
Breakfast	0.01	0.000	0.02	1.10	0.02	0.16	0.01
Olive	0.01	0.000	0.02	1.15	0.04	0.19	0.01
Cheese	0.00	0.000	0.01	0.42	0.00	0.06	0.00
Sugar	0.01	0.000	0.01	0.75	0.01	0.10	0.01
Salt and Spices	0.03	0.001	0.03	2.54	0.12	0.35	0.02
Bakery and Patisserie	0.01	0.000	0.01	0.85	0.01	0.11	0.01
Liquid Oil	0.01	0.000	0.01	0.62	0.01	0.08	0.00
White Meat	0.04	0.003	0.06	3.95	0.31	0.56	0.03
Meat Deli	0.01	0.000	0.02	1.12	0.02	0.15	0.01
Fish Seafood	0.01	0.000	0.01	0.69	0.01	0.11	0.01
Oral Care	0.03	0.002	0.05	3.44	0.22	0.47	0.02
Washing up	0.06	0.007	0.08	6.22	0.65	0.80	0.04
Laundry Utensils	0.01	0.000	0.02	1.31	0.03	0.17	0.01
Washing	0.02	0.001	0.03	1.76	0.07	0.26	0.01
Shaving Supplies	0.03	0.002	0.05	2.82	0.20	0.45	0.02

The forecasting performance of the Generalized Additive Model is presented in Table 17. The RSquared values for the GAMs range from 0.4611 to 0.9581, indicating a diverse range of predictive performance across different product categories. Adjusted R-2 values generally follow the same trend. Hit Ratio and DS values show the ability of the model to predict outcomes accurately and discriminate between different categories. These metrics provide insights into the accuracy and precision of the GAMs. Lower values are generally desirable. UTheil coefficients provide a measure of forecasting accuracy relative to a benchmark model. Smaller values suggest better forecasting performance. "Sugar" and "Mineral Water" categories exhibit high RSquared values, indicating effective model performance. "Washing up" category shows a relatively low RSquared value, suggesting that the model may not be as effective for predicting sales in this category. As a result, the Generalized Additive Models show diverse performance across different product categories. The choice of the appropriate model for each category should consider the specific characteristics and patterns observed in the data. These insights provide a comprehensive understanding of the Generalized Additive Models' performance in predicting sales for various product categories. It's essential to consider the specific context of each category when interpreting the model's predictive capabilities.

Table 17
Performance of Generalized Additive Model

	R^2	Adj R^2	Hit Ratio	DS	MAE	MSE	RMSE
Carbonated Beverage	0.65	0.65	0.55	0.52	0.09	0.02	0.12
Still Drink	0.68	0.68	0.62	0.60	0.08	0.01	0.11
Mineral Water	0.69	0.69	0.60	0.54	0.14	0.03	0.18
Water	0.73	0.72	0.60	0.55	0.06	0.01	0.08
Milk	0.73	0.73	0.62	0.60	0.11	0.02	0.15
Instant Meal	0.76	0.76	0.71	0.61	0.09	0.01	0.12
Frozen Food	0.74	0.74	0.60	0.56	0.22	0.08	0.29
Vegetables	0.74	0.73	0.64	0.61	0.25	0.14	0.37
Fruits	0.42	0.41	0.57	0.55	0.75	0.93	0.96
Yogurt	0.71	0.71	0.56	0.55	0.16	0.04	0.20
Egg	0.75	0.75	0.58	0.42	0.25	0.09	0.31
Breakfast	0.46	0.46	0.64	0.58	0.33	0.18	0.43
Olive	0.73	0.73	0.54	0.52	0.40	0.37	0.61
Cheese	0.61	0.61	0.58	0.57	0.30	0.14	0.38
Sugar	0.96	0.96	0.54	0.29	0.09	0.02	0.13
Salt and Spices	0.85	0.85	0.53	0.49	0.44	0.30	0.55
Bakery and Patisserie	0.62	0.61	0.63	0.61	0.12	0.02	0.15
Liquid Oil	0.77	0.77	0.54	0.49	0.56	0.57	0.75
White Meat	0.54	0.54	0.56	0.52	0.99	1.51	1.23
Meat Deli	0.61	0.61	0.61	0.56	1.22	2.52	1.59
Fish Seafood	0.59	0.59	0.55	0.47	0.74	0.84	0.92
Oral Care	0.84	0.84	0.55	0.51	1.38	3.62	1.90
Washing up	0.58	0.57	0.56	0.53	1.07	1.92	1.38
Laundry Utensils	0.80	0.80	0.48	0.34	0.73	0.95	0.97
Washing	0.69	0.69	0.62	0.60	0.46	0.50	0.71
Shaving Supplies	0.66	0.65	0.58	0.57	1.66	6.35	2.52
	MARE	MSRE	RMSRE	MAPE	MSPE	RMSPE	UTheil
Carbonated Beverage	0.01	0.000	0.02	1.18	0.02	0.15	0.01
Still Drink	0.01	0.000	0.02	1.15	0.02	0.15	0.01
Mineral Water	0.02	0.001	0.02	1.65	0.05	0.22	0.01
Water	0.01	0.000	0.02	1.42	0.03	0.18	0.01
Milk	0.01	0.000	0.02	1.16	0.02	0.16	0.01
Instant Meal	0.01	0.000	0.02	1.29	0.03	0.17	0.01
Frozen Food	0.01	0.000	0.02	1.24	0.03	0.17	0.01
Vegetables	0.03	0.002	0.04	2.52	0.16	0.40	0.02
Fruits	0.05	0.005	0.07	5.07	0.47	0.68	0.03
Yogurt	0.02	0.000	0.02	1.60	0.04	0.20	0.01
Egg	0.01	0.000	0.02	1.42	0.03	0.18	0.01
Breakfast	0.01	0.000	0.02	1.49	0.04	0.19	0.01
Olive	0.01	0.001	0.02	1.48	0.05	0.23	0.01
Cheese	0.01	0.000	0.01	0.83	0.01	0.11	0.01
Sugar	0.01	0.000	0.01	0.86	0.01	0.12	0.01
Salt and Spices	0.04	0.002	0.05	3.80	0.24	0.49	0.02
Bakery and Patisserie	0.01	0.000	0.01	1.15	0.02	0.15	0.01
Liquid Oil	0.01	0.000	0.01	1.01	0.02	0.13	0.01
White Meat	0.04	0.002	0.05	3.97	0.24	0.49	0.02
Meat Deli	0.02	0.001	0.03	2.26	0.08	0.29	0.02
Fish Seafood	0.01	0.000	0.02	1.42	0.03	0.18	0.01

Table 17
Continued

	R^2	Adj R^2	Hit Ratio	DS	MAE	MSE	RMSE
Oral Care	0.03	0.002	0.05	3.45	0.22	0.47	0.02
Washing up	0.04	0.002	0.05	3.60	0.22	0.47	0.02
Laundry Utensils	0.01	0.000	0.02	1.16	0.02	0.15	0.01
Washing	0.02	0.001	0.03	1.86	0.08	0.29	0.01
Shaving Supplies	0.03	0.003	0.05	3.28	0.25	0.50	0.03

4.5. Key observations

Key observations from this research can be summarized as follows:

x. Linear Regression Models: Linear regression models applied to performance metrics characterized by “smaller-is-better” traits exhibit notably higher r^2 values when examined in conjunction with standard deviations of the product categories. This suggests a more pronounced linear relationship between standard deviation and these metrics compared to those with “higher-is-better” characteristics (Figure 12).

xi. Hit Ratio and DS: Hit ratio and DS metrics display a minimal linear association with standard deviation, indicating that the correlation between these metrics and standard deviation is weak or negligible (Figure 12).

xii. EMD Impact: The integration of Empirical Mode Decomposition (EMD) into the single Auto Regressive Integrated Moving Average (ARIMA) model substantially enhances its performance, with this improvement holding statistical significance (Figure 11). This finding aligns with similar observations in the existing literature.

xiii. Optimal p-q Distribution: The distribution of optimal p-q values exhibits a right-skewed shape, indicating that a small number of degrees suffice for modeling the sub-components of the time series (Table 10).

xiv. Maximum Difference Number: It was determined that a maximum difference of four times is required to render all subcomponents stationary, reflecting the stability of the data transformation process (Table 10).

xv. Rolling Window Forecasting: Analysis of the coefficients in the rolling window forecasting scheme reveals that the constant coefficient remains relatively consistent across the 90 ARIMA models, while other parameters exhibit slight variations (Figure 6).

The ARIMA and machine learning algorithms has facilitated the attainment of diverse performance metrics across a broad spectrum. It is challenging to unequivocally assert the superiority of one technique over the other under all circumstances. This complexity arises from the nuanced nature of the datasets and the distinct strengths exhibited by each method in various contexts. ARIMA is adept at capturing temporal dependencies and trends within time series data. Its strength lies in modelling linear relationships and exhibiting robust performance in scenarios where such relationships dominate. On the other hand, machine learning algorithms, characterized by their flexibility and adaptability, excel in capturing non-linear patterns and complex dependencies present in the data.

5. Conclusion

In this comprehensive research, we have introduced a sophisticated price forecasting system, amalgamating Empirical Mode Decomposition (EMD) and Auto Regressive Integrated Moving Average (ARIMA) models, to forecast retail category prices for an online retail enterprise operating in the Turkish market. Our 900-day dataset, meticulously gathered through daily web scraping, has formed the bedrock of our research. We conducted a thorough evaluation of the forecasting performance using a diverse set of fourteen metrics. The results, corroborated by the Wilcoxon signed-rank test, decisively establish the statistical superiority of our hybrid model over the singular ARIMA model, as manifested in the performance metrics. The forecasting process is repeated for machine learning algorithms namely, Neural Networks, Support Vector Regression, Regression Tree, Gaussian Process Regression, and Generalized Additive Model.

It is imperative to recognize that the efficacy of each technique is contingent upon the specific characteristics and inherent complexities of the dataset at hand. No single method universally outshines the other across all conceivable scenarios. This acknowledgment underscores the importance of a nuanced and context-specific evaluation when selecting an analytical approach. It is in this appreciation of the nuanced interplay between methodologies that a more informed and judicious choice can be made, tailored to the intricacies of the dataset and the analytical objectives.

Linear regression analysis is utilized to examine the relationship between category standard deviations and forecasting performance. Furthermore, our analysis ventured into the intriguing relationship between category standard deviations and forecasting performance, shedding light on a prominent trend: categories characterized by lower standard deviations consistently demonstrated heightened forecasting accuracy. This underscores the pivotal role of price stability in forecasting precision.

Nonetheless, we acknowledge certain limitations inherent in our study, particularly related to the temporal constraints of our dataset collection. Given the recent high inflation levels experienced within the Turkish economy, which could potentially compromise the integrity of the collected dataset and diminish the forecasting efficacy of our model, we took a deliberate approach in gathering data during a relatively more economically stable period characterized by reduced price volatility.

The implications of our research resonate beyond the realm of online retail, extending to the broader supply chain and business landscape. As data-driven insights become increasingly instrumental in decision-making, our work illuminates a pathway to more accurate retail price forecasting, which can be harnessed by retailers to refine their promotional strategies and ultimately benefit both businesses and consumers.

In this paper, the parameters of the ARIMA are determined before forecasting the testing set, and then along the testing period, the parameters are fixed. Future studies may explore avenues for dynamic ARIMA parameter determination, advanced EMD (EEMD, CEEMDAN, ICEEMDAN) variants, alternative modelling approaches like machine learning, and the application of our hybrid model to diverse time series, offering exciting prospects for further research and practical implementation in various domains.

Acknowledgement

This research was not funded by any grant.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Hoeltgebaum, H., Borenstein, D., Fernandes, C., & Veiga, Á. (2021). A score-driven model of short-term demand forecasting for retail distribution centers. Journal of Retailing, 97(4), 715-725. https://doi.org/10.1016/j.jretai.2021.05.003

[2] Wang, J., Chong, W. K., Lin, J., & Hedenstierna, C. P. T. (2024). Retail Demand Forecasting Using Spatial-Temporal Gradient Boosting Methods. Journal of Computer Information Systems, 64(5), 652-664. https://doi.org/10.1080/08874417.2023.2240753

[3] Fedoseeva, S., & Herrmann, R. (2023). Assortments and prices in online grocery retailing. Digital Business, 3(1), 100054. https://doi.org/10.1016/j.digbus.2023.100054

[4] Sharma, R. R., Kumar, M., Maheshwari, S., & Ray, K. P. (2020). EVDHM-ARIMA-based time series forecasting model and its application for COVID-19 cases. IEEE Transactions on Instrumentation and Measurement, 70, 1-10. https://doi.org/10.1109/tim.2020.3041833

[5] Lasheras, F. S., de Cos Juez, F. J., Sánchez, A. S., Krzemień, A., & Fernández, P. R. (2015). Forecasting the COMEX copper spot price by means of neural networks and ARIMA models. Resources Policy, 45, 37-43. https://doi.org/10.1016/j.resourpol.2015.03.004

[6] Qin, Q., Huang, Z., Zhou, Z., Chen, Y., & Zhao, W. (2022). Hodrick–Prescott filter-based hybrid ARIMA–SLFNs model with residual decomposition scheme for carbon price forecasting. Applied Soft Computing, 119, 108560. https://doi.org/10.1016/j.asoc.2022.108560

[7] Zhu, B., & Wei, Y. (2013). Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. Omega, 41(3), 517–524. https://doi.org/10.1016/j.omega.2012.06.005

[8] Liu, J., Wang, P., Chen, H., & Zhu, J. (2022). A combination forecasting model based on hybrid interval multi-scale decomposition: Application to interval-valued carbon price forecasting. Expert Systems with Applications, 191, 116267. https://doi.org/10.1016/j.eswa.2021.116267

[9] Matyjaszek, M., Fernández, P. R., Krzemień, A., Wodarski, K., & Valverde, G. F. (2019). Forecasting coking coal prices by means of ARIMA models and neural networks, considering the transgenic time series theory. Resources Policy, 61, 283-292. https://doi.org/10.1016/j.resourpol.2019.02.017

[10] Alam, M. S., Murshed, M., Manigandan, P., Pachiyappan, D., & Abduvaxitovna, S. Z. (2023). Forecasting oil, coal, and natural gas prices in the pre-and post-COVID scenarios: contextual evidence from India using time series forecasting tools. Resources Policy, 81, 103342. https://doi.org/10.1016/j.resourpol.2023.103342

[11] Kriechbaumer, T., Angus, A., Parsons, D., & Casado, M. R. (2014). An improved wavelet–ARIMA approach for forecasting metal prices. Resources Policy, 39, 32-41. https://doi.org/10.1016/j.resourpol.2013.10.005

[12] Lehna, M., Scheller, F., & Herwartz, H. (2022). Forecasting day-ahead electricity prices: A comparison of time series and neural network models taking external regressors into account. Energy Economics, 106, 105742. https://doi.org/10.1016/j.eneco.2021.105742

[13] Zhang, B., Song, C., Jiang, X., & Li, Y. (2023). Electricity price forecast based on the STL-TCN-NBEATS model. Heliyon, 9(1), e13029. https://doi.org/10.1016/j.heliyon.2023.e13029

[14] Krishna Prakash, N., & Singh, J. G. (2023). Electricity price forecasting using hybrid deep learned networks. Journal of Forecasting, 42(7), 1750-1771. https://doi.org/10.1002/for.2981

[15] Mohammadi, H., & Su, L. (2010). International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. Energy Economics, 32(5), 1001-1008. https://doi.org/10.1016/j.eneco.2010.04.009

[16] Jianwei, E., Bao, Y., & Ye, J. (2017). Crude oil price analysis and forecasting based on variational mode decomposition and independent component analysis. Physica A: Statistical Mechanics and its Applications, 484, 412-427. https://doi.org/10.1016/j.physa.2017.04.160

[17] Koutroumanidis, T., Ioannou, K., & Arabatzis, G. (2009). Predicting fuelwood prices in Greece with the use of ARIMA models, artificial neural networks and a hybrid ARIMA–ANN model. Energy Policy, 37(9), 3627-3634. https://doi.org/10.1016/j.enpol.2009.04.024

[18] Pai, P. F., & Lin, C. S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. Omega, 33(6), 497-505. https://doi.org/10.1016/j.omega.2004.07.024

[19] Zolfaghari, M., & Gholami, S. (2021). A hybrid approach of adaptive wavelet transform, long short-term memory and ARIMA-GARCH family models for the stock index prediction. Expert Systems with Applications, 182, 115149. https://doi.org/10.1016/j.eswa.2021.115149

[20] Li, C., & Hu, J. W. (2012). A new ARIMA-based neuro-fuzzy approach and swarm intelligence for time series forecasting. Engineering Applications of Artificial Intelligence, 25(2), 295-308. https://doi.org/10.1016/j.engappai.2011.10.005

[21] Babu, C. N., & Reddy, B. E. (2015). Prediction of selected Indian stock using a partitioning–interpolation based ARIMA–GARCH model. Applied Computing and Informatics, 11(2), 130-143. https://doi.org/10.1016/j.aci.2014.09.002

[22] Yu, Z., Qin, L., Chen, Y., & Parmar, M. D. (2020). Stock price forecasting based on LLE-BP neural network model. Physica A: Statistical Mechanics and Its Applications, 553, 124197. https://doi.org/10.1016/j.physa.2020.124197

[23] David, S. A., Inacio Jr, C. M. C., Nunes, R., & Machado, J. T. (2021). Fractional and fractal processes applied to cryptocurrencies price series. Journal of Advanced Research, 32, 85-98. https://doi.org/10.1016/j.jare.2020.12.012

[24] Büyükşahin, Ü. Ç., & Ertekin, Ş. (2019). Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. Neurocomputing, 361, 151-163. https://doi.org/10.1016/j.neucom.2019.05.099

[25] Bjørnland, H. C., Gerdrup, K., Jore, A. S., Smith, C., & Thorsrud, L. A. (2012). Does forecast combination improve Norges Bank inflation forecasts?. Oxford Bulletin of Economics and Statistics, 74(2), 163-179. https://doi.org/10.1111/j.1468-0084.2011.00639.x

[26] Bataa, E., Osborn, D. R., Sensier, M., & Dijk, D. V. (2014). Identifying changes in mean, seasonality, persistence and volatility for G7 and Euro Area inflation. Oxford Bulletin of Economics and Statistics, 76(3), 360-388. https://doi.org/10.1111/obes.12021

[27] Banaś, J., & Utnik-Banaś, K. (2021). Evaluating a seasonal autoregressive moving average model with an exogenous variable for short-term timber price forecasting. Forest Policy and Economics, 131, 102564. https://doi.org/10.1016/j.forpol.2021.102564

[28] An, W., Wang, L., & Zhang, D. (2023). Comprehensive commodity price forecasting framework using text mining methods. Journal of Forecasting, 42(7), 1865-1888. https://doi.org/10.1002/for.2985

[29] An, W., Wang, L., & Zeng, Y. R. (2023). Text-based soybean futures price forecasting: A two-stage deep learning approach. Journal of Forecasting, 42(2), 312-330. https://doi.org/10.1002/for.2909

[30] Ong, C. S., Huang, J. J., & Tzeng, G. H. (2005). Model identification of ARIMA family using genetic algorithms. Applied Mathematics and Computation, 164(3), 885-912. https://doi.org/10.1016/j.amc.2004.06.044

[31] Madziwa, L., Pillalamarry, M., & Chatterjee, S. (2022). Gold price forecasting using multivariate stochastic model. Resources Policy, 76, 102544. https://doi.org/10.1016/j.resourpol.2021.102544

[32] Deina, C., do Amaral Prates, M. H., Alves, C. H. R., Martins, M. S. R., Trojan, F., Stevan Jr, S. L., & Siqueira, H. V. (2022). A methodology for coffee price forecasting based on extreme learning machines. Information Processing in Agriculture, 9(4), 556-565. https://doi.org/10.1016/j.inpa.2021.07.003

[33] Pedregal, D. J. (2020). Forecasting uranium prices: Some empirical results. Nuclear Engineering and Technology, 52(6), 1334-1339. https://doi.org/10.1016/j.net.2019.11.028

[34] Bloznelis, D. (2018). Short-term salmon price forecasting. Journal of Forecasting, 37(2), 151-169. https://doi.org/10.1002/for.2482

[35] Imbat, M. B. S., Gordovez, F. S. S., Solomon, R. M., Malicdem, G. E. Q., Andrada, J. D. P., & Belandres, E. B. (2024). Rice Price Forecasting Using the Arima Model, Iconic Research and Engineering Journals, 7(7), 333-338.

[36] Vo, N., & Ślepaczuk, R. (2022). Applying hybrid ARIMA-SGARCH in algorithmic investment strategies on S&P500 index. Entropy, 24(2), 158. https://doi.org/10.3390/e24020158

[37] Wolters, J., & Huchzermeier, A. (2021). Joint in-season and out-of-season promotion demand forecasting in a retail environment. Journal of Retailing, 97(4), 726-745. https://doi.org/10.1016/j.jretai.2021.01.003

[38] Sarlo, R., Fernandes, C., & Borenstein, D. (2023). Lumpy and intermittent retail demand forecasts with score-driven models. European Journal of Operational Research, 307(3), 1146-1160. https://doi.org/10.1016/j.ejor.2022.10.006

[39] Li, C., & Lim, A. (2018). A greedy aggregation–decomposition method for intermittent demand forecasting in fashion retailing. European Journal of Operational Research, 269(3), 860-869. https://doi.org/10.1016/j.ejor.2018.02.029

[40] Ma, S., Fildes, R., & Huang, T. (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra-and inter-category promotional information. European Journal of Operational Research, 249(1), 245-257. https://doi.org/10.1016/j.ejor.2015.08.029

[41] McIntyre, S. H., Achabal, D. D., & Miller, C. M. (1993). Applying case-based reasoning to forecasting retail sales. Journal of Retailing, 69(4), 372-398. https://doi.org/10.1016/0022-4359(93)90014-a

[42] 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(2), 523-539. https://doi.org/10.1016/j.ejor.2023.10.039

[43] Ma, S., & Fildes, R. (2021). Retail sales forecasting with meta-learning. European Journal of Operational Research, 288(1), 111-128. https://doi.org/10.1016/j.ejor.2020.05.038

[44] Huang, T., Fildes, R., & Soopramanien, D. (2014). The value of competitive information in forecasting FMCG retail product sales and the variable selection problem. European Journal of Operational Research, 237(2), 738-748. https://doi.org/10.1016/j.ejor.2014.02.022

[45] Schlaich, T., & Hoberg, K. (2024). When is the next order? Nowcasting channel inventories with Point-of-Sales data to predict the timing of retail orders. European Journal of Operational Research, 315(1), 35-49. https://doi.org/10.1016/j.ejor.2023.10.038

[46] Agrawal, D., & Schorling, C. (1996). Market share forecasting: An empirical comparison of artificial neural networks and multinomial logit model. Journal of Retailing, 72(4), 383-407. https://doi.org/10.1016/s0022-4359(96)90020-2

[47] Rahman, M. A., Modak, C., Mozumder, M. A. S., Miah, M. N. I., Hasan, M., Sweet, M. M. R., Hossan, M.Z., & Alam, M. (2024). Advancements in Retail Price Optimization: Leveraging Machine Learning Models for Profitability and Competitiveness. Journal of Business and Management Studies, 6(3), 103-110. https://doi.org/10.32996/jbms.2024.6.3.11

[48] Box, G.E.P., Jenkins, G.M., Reinsel, G.C. & Ljung, G.M. (2016) Time Series Analysis: Forecasting and Control. Fifth Edition, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken.

[49] Liu, H., Tian, H. Q., & Li, Y. F. (2015). An EMD-recursive ARIMA method to predict wind speed for railway strong wind warning system. Journal of Wind Engineering and Industrial Aerodynamics, 141, 27-38. https://doi.org/10.1016/j.jweia.2015.02.004

[50] Wang, Z. Y., Qiu, J., & Li, F. F. (2018). Hybrid models combining EMD/EEMD and ARIMA for long-term streamflow forecasting. Water, 10(7), 853. https://doi.org/10.3390/w10070853

[51] Li, T., Qu, S., & Huang, G. (2021). Research on the prediction of Shenzhen growth enterprise market price index based on EMD-ARIMA model. In LISS 2020: Proceedings of the 10th International Conference on Logistics, Informatics and Service Sciences (pp. 783-795). Springer Singapore. https://doi.org/10.1007/978-981-33-4359-7_54

[52] Fatema, N., Malik, H., & Abd Halim, M. S. (2022). Hybrid approach combining EMD, ARIMA and monte carlo for multi-step ahead medical tourism forecasting. Journal of Intelligent & Fuzzy Systems, 42(2), 1235–1251. https://doi.org/10.3233/jifs-189785

[53] Powell, B., Nason, G., Elliott, D., Mayhew, M., Davies, J., & Winton, J. (2018). Tracking and Modelling Prices Using Web-Scraped Price Microdata: Towards Automated Daily Consumer Price Index Forecasting. Journal of the Royal Statistical Society Series A: Statistics in Society, 181(3), 737–756. https://doi.org/10.1111/rssa.12314

[54] Claveria, O., Monte, E., & Torra, S. (2017). Using survey data to forecast real activity with evolutionary algorithms. A cross-country analysis. Journal of Applied Economics, 20(2), 329–349. https://doi.org/10.1016/s1514-0326(17)30015-6

[55] Liu, M. D., Ding, L., & Bai, Y. L. (2021). Application of hybrid model based on empirical mode decomposition, novel recurrent neural networks and the ARIMA to wind speed prediction. Energy Conversion and Management, 233, 113917. https://doi.org/10.1016/j.enconman.2021.113917

[56] Cao, Y., Zhang, D., Ding, S., Zhong, W., & Yan, C. (2024). A Hybrid Air Quality Prediction Model Based on Empirical Mode Decomposition. Tsinghua Science and Technology, 29(1), 99–111. https://doi.org/10.26599/tst.2022.9010060

[57] Wang, H., Liu, L., Dong, S., Qian, Z., & Wei, H. (2015). A novel work zone short-term vehicle-type specific traffic speed prediction model through the hybrid EMD–ARIMA framework. Transportmetrica B: Transport Dynamics, 4(3), 159–186. https://doi.org/10.1080/21680566.2015.1060582

[58] Abadan, S., & Shabri, A. (2014). Hybrid Empirical Mode Decomposition-ARIMA for Forecasting Price of Rice. Applied Mathematical Sciences, 8(63), 3133–3143. https://doi.org/10.12988/ams.2014.43189

[59] Nasir, J., Aamir, M., Haq, Z. U., Khan, S., Amin, M. Y., & Naeem, M. (2023). A New Approach for Forecasting Crude Oil Prices Based on Stochastic and Deterministic Influences of LMD Using ARIMA and LSTM Models. IEEE Access, 11, 14322–14339. https://doi.org/10.1109/access.2023.3243232

[60] Yang, H. L., & Lin, H. C. (2016). An integrated model combined ARIMA, EMD with SVR for stock indices forecasting. International Journal on Artificial Intelligence Tools, 25(02), 1650005. https://doi.org/10.1142/s0218213016500056

[61] Zhou, Y., & Huang, M. (2016). Lithium-ion batteries remaining useful life prediction based on a mixture of empirical mode decomposition and ARIMA model. Microelectronics Reliability, 65, 265–273. https://doi.org/10.1016/j.microrel.2016.07.151

[62] Awajan, A. M., Al-Hasanat, B., Elkaroui, E., AL e'damat, A., Al-Gounmeein, R. S., Al-Jawarneh, A. S., Ayyoub H. N., & Alfarajat, E. (2024). Time Series Forecasting of New Cases for COVID-19 Pandemic in Jordan Using Enhanced Hybrid EMD-ARIMA. Journal of Statistics Applications & Probability An International Journal, 13(1), 261. https://doi.org/10.18576/jsap/130118

[63] Chyon, F. A., Suman, M. N. H., Fahim, M. R. I., & Ahmmed, M. S. (2022). Time series analysis and predicting COVID-19 affected patients by ARIMA model using machine learning. Journal of Virological Methods, 301, 114433. https://doi.org/10.1016/j.jviromet.2021.114433

[64] Faruk, D. Ö. (2010). A hybrid neural network and ARIMA model for water quality time series prediction. Engineering Applications of Artificial Intelligence, 23(4), 586–594. https://doi.org/10.1016/j.engappai.2009.09.015

[65] PennState. (2023). https://online.stat.psu.edu/stat510/lesson/2/2.2

[66] Duke. (2023). https://people.duke.edu/~rnau/arimrule.htm

[67] Mohammed, E. A., Naugler, C., & Behrouz, H. (2015). Emerging Business Intelligence Framework for a Clinical Laboratory Through Big Data Analytics. Emerging Trends in Computational Biology, Bioinformatics, and Systems Biology: Algorithms and Software Tools, 577–602. https://doi.org/10.1016/b978-0-12-802508-6.00032-6

[68] Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N. C., Tung, C. C., & Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 454(1971), 903–995. https://doi.org/10.1098/rspa.1998.0193

[69] Lin, Y., Yan, Y., Xu, J., Liao, Y., & Ma, F. (2021). Forecasting stock index price using the CEEMDAN-LSTM model. The North American Journal of Economics and Finance, 57, 101421. https://doi.org/10.1016/j.najef.2021.101421

[70] Zhang, Y., Yan, B., & Aasma, M. (2020). A novel deep learning framework: Prediction and analysis of financial time series using CEEMD and LSTM. Expert Systems with Applications, 159, 113609. https://doi.org/10.1016/j.eswa.2020.113609

[71] Ning, Y., Kazemi, H., & Tahmasebi, P. (2022). A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet. Computers & Geosciences, 164, 105126. https://doi.org/10.1016/j.cageo.2022.105126

[72] Matthews, K. (1985). Forecasting with a Rational Expectations Model of The UK. Oxford Bulletin of Economics and Statistics, 47(4), 311–336. https://doi.org/10.1111/j.1468-0084.1985.mp47004001.x

[73] Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing, 415, 295–316. https://doi.org/10.1016/j.neucom.2020.07.061