Advances in Industrial Engineering, Autumn 2019, 53(4): 169-184
DOI: 10.22059/jieng.2021.316849.1746

RESEARCH PAPER

A Novel Approach for Multi Product Demand Forecast Using Data Mining Techniques (Empirical Study: Carpet Industry)

Sayed Mohammad-Reza Vaghefinezhad [a,*], Jafar Razmi[b], Fariborz Jolai[b]

[a] *Department of Industrial Engineering, Kish International Campus, University of Tehran, Iran*
[b] *School of Industrial Engineering, College of Engineering, University of Tehran, Iran*

Received: 07 January 2021, Revised: 21 January 2021, Accepted: 21 January 2021

Abstract

Accurate demand forecasting plays an important role in meeting customers' expectations and satisfaction that strengthen the enterprise's competitive position. In this research, time series and artificial neural networks methods compete to provide more precise demand estimation while having a large variety of products. After obtaining the initial results, suggestions have been implemented to improve forecasting accuracy. As a direct result of that, the average mean absolute percentage error (MAPE) of all products' demand forecast reduces significantly. To improve the quality of historical records, association rules and substitution ratio have been applied. This method plays a significant role to detect the existing pattern in historical data and MAPE reduction. The satisfactory and applicable results provide the company with a more accurate forecast. Moreover, the issue of precepting confusing historical data which caused unforecastable trends has been solved. The R language and "neuralnet", "nnfor", "forecast", and "arules" packages have been applied in programming.

Keywords:
Artificial Neural Network;
Association Rules;
Demand Forecasting;
Data Mining;
R Language;
Time Series

Introduction

Accurate demand estimation contributes significantly to the different stages of production from supply to products sale. In the real world, there are many products with different behaviors in the production company portfolio. Various products while each of them may have a different stage of the life cycle. So, a proper forecasting method is needed to identify systematic forecasting factors such as level, trend, and seasonality or learn the trend for each of them [1].

Moreover, each product during each stage of its life cycle may include only one or all level, trend, and seasonality components. In other words, several models or settings of smoothing constant may require to predict demand properly. For instance, a product in the growth stage can include level and trend, while the same product in the maturation cycle can only contain level or level and seasonality [2]. Hence, the application of various methods or settings is essential not only to predict different products but also to detect the behavior of a product at different stages of its life cycle. According to the above, relying on one method to estimate the demand for all products or a product in the whole of its life cycle will reduce the overall accuracy.

Historical data is typically used as an input to prediction models. There are several concerns with using this past data. As an instance, data may show a particular tendency and behavior due

[*] Corresponding author: (S. M. R. Vaghefinezhad)
Email: jrazmi@ut.ac.ir

to a particular event in the past. In this case, even with the best combination of methods and techniques, there will be a significant error in prediction. Accurate prediction plays a key role in the appropriate and timely process of supply, manufacturing, and storing of marketable products. Reviewing the related literature reveals that most of the time historical data of one product is used for forecasting or find the more accurate techniques or settings while in real cases there are multiple products in the portfolio of the companies. Furthermore, only outlier data will be omitted to increase the precision of prediction. Finding the cause of these patterns that make the historical data unpredictable has less been examined.

In this study, a novel and comprehensive approach has been proposed to deal with multi-product and product family prediction for practitioners. Time series and neural networks were used to estimate demand with the least mean absolute percentage error (MAPE). MAPE error was used to comparing largely because overestimate and underestimate were both undesirable. In addition, this kind of error is comparable for all products. Sales` records of 511 products during the 4 years of sales from 2016 to 2019 were extracted from the Oracle database of the organization with separate purchase requests. Data was transferred to Excel and saved in Comma Separated Values (CSV) format to create a report of sales by aggregating data on a monthly basis and transactional data. This data was analyzed in R software version 4.03 using “neuralnet”, “nnfor”, “forecast” and “arules” packages. The optimal method for predicting each product was selected based on the MAPE. Then, the forecast for next year for each product was completed by the optimal method. Cumulative sales of products were also forecasted.

It is also required to predict product families in aggregate planning. Thus, the optimum method of forecasting the known product families in the organization was also presented. The possibility of replacing the order from one product to another similar product while inventory shortage exists. This issue affected historical data and make them unforecastable. So, substitute products were identified by exploring the association rules in the transactional database. By considering the same usage, the identical explored substitute products were integrated and entered into the predicted model. The error of the new groups was reduced compared to the conventional ones. Then, ABX-XYZ analysis is applied in new groups to categorized substitute products in each group. This step helps to find which products must be substituted in favor of the superior ones. Then suggestions were offered with the aim of reducing sales fluctuations and enhancing the sales of dominant products.

The contributions of this research can be named in the following order:

- Using association rules to find substitute products in a large volume of data and perceive confusing historical data trend
- Comparing methods to predict each product while there is multi- product
- Selection of superior substitute based on ABC-XYZ method
- Empirical implementation of the proposed model for studies while most similar studies used numerical examples.

Literature Review

Demand forecasting plays a leading role in scheduling decision making. Imprecise demand estimation, on the one hand, raises the probability of backlog, and on the other hand, can be resulted in overproduction and inventory costs. Although the role of forecasting seems to be more significant in make-to-stock strategy, its absence causes a long procurement process in just-in-time and make-to-order systems. Data mining techniques have been widely applied for patterns discovering in data and prediction [3][4]. Over the last seven decades, a large and growing body of literature has investigated forecasting techniques. Chopra and Meindl [5] stated that previous observed demands consisted of systematic and random components. The systematic component is defined as the demand expected value. The random component cannot be forecasted. Nevertheless, its variability and size can be estimated by measuring the error of the forecasting technique. The systematic component generally can be composed by level, trend, and seasonality. However, a major problem with forecasting is seasonality. Forecasting techniques are classified as qualitative, time-series, causal, and simulation [5][6]. The qualitative methods rely on experts' judgment while time-series try to estimate demand using historical data. The causal techniques develop a prediction by finding a relation between demand and some other factors. Simulation techniques try to mimic the demand by combining causal and time series methods [5].

Time series methods

Time-series techniques are divided into static and adaptive. In adaptive approach, level, trend, and seasonality will be updated whenever new demand is recorded [7][8][9]. A useful classification recognized five and three types of trend and seasonality; respectively.

Different kinds of trend are: without trend, additive trend, damped additive trend, multiplicative trend, damped multiplicative trend. Seasonality was broken into: without seasonality, additive and multiplicative seasonality [10][11][1].

Hence, existing forecasting approaches can be organized in 15 various classes (Table 1 and 2). For instance, simple exponential smoothing has been utilized while there is neither trend nor seasonality [5][12]. Holt's technique models additive trend with or without seasonality. Holt–Winters' additive and multiplicative methods forecast demand with additive and multiplicative trend and seasonality, respectively [5][13][14]. The multiplicative attribute shows the dependency of trend and seasonality with the level whereas the additive characteristic reveals that trend and seasonality are not affected by level in the systematic component [15]. Chopra and Meindl [5] added mixed state and reclassified systematic component of observed data to multiplicative, additive, and mixed shown in Eq. 1.

$$
\begin{aligned}
&\text{Multiplicative:} && \textit{Level} * \textit{Trend} * \textit{Seasonality}\\
&\text{Additive:} && \textit{Level} + \textit{Trend} + \textit{Seasonality}\\
&\text{Mixed:} && (\textit{Level} + \textit{Trend}) * \textit{Seasonality}
\end{aligned}
\qquad (1)
$$

Table 1. Different groups of forecast methods based on a variety of trend and seasonality [1]

Trend Component	Seasonal Component N (None)	A (Additive)	M (Multiplicative)
N (None)	N,N	N,A	N,M
A (Additive)	A,N	A,A	A,M
A_d (Additive damped)	A_d,N	A_d, A	A_d, M
M (Multiplicative)	M, N	M, A	M, M
M_d (Multiplicative damped)	M_d ,N	M_d ,A	M_d,M

(N,N) : simple exponential smoothing method
(A,N) : Holt's linear method
(A_d,N) : damped trend method
(A,A) : Holt-Winters' additive method
(A,M) : Holt-Winters' multiplicative method

Two types of deviations have been distinguished in seasonal time series. These variations exist between the two following data and between the averages of successive seasons.

Seasonality is particularly recognizable in the economic and business time series. Time series models generally support trends and seasonality. Various techniques have been developed for trend and seasonality modeling in the literature. Some of the well-known approaches are simple seasonal exponential smoothing, Box–Jenkins, Holt-Winter's, and artificial neural networks (ANNs) [15].

Simple seasonal exponential smoothing technique cannot support trend and is applied as time series only include level and seasonality. Another well-known forecasting model is Box–Jenkins or Autoregressive integrated moving average (ARIMA) [16].

Seasonal ARIMA that consists of seasonal component is known as SARIMA. This model has been confined with the linearity assumption and is unsuitable for non-linear problems [17]. Holt- Winter's model is a seasonal extension of the exponential smoothing technique. This method is compared with Box–Jenkins by considering these factors: effectiveness, adaptiveness, simply using and required input data [18].

Parameters

α	level smoothing constant	$(0 \leq \alpha \leq 1)$
β	trend smoothing constant	$(0 \leq \beta \leq 1)$
λ	seasonality smoothing constant	$(0 \leq \lambda \leq 1)$
ϕ	damped constant	$(0 \leq \phi \leq 1)$

$$\phi_h = \phi + \phi_2 + \cdots + \phi_h \tag{2}$$

ε_t	forecast error of period t
ℓ_t	level for period t
b_t	trend for period t
S_t	trend for period t
$\hat{y}_{t+h\|t}$	forecasted demand of h^{th} period after the last historical data
m	number of seasons
θ	constant of moving average
ϕ	constant of autoregression
B	backshift notation, it means the demand of previous period

Exponential smoothing and ARIMA are widely used in time series. While in exponential smoothing methods, data description is performed based on trends and seasonality, ARIMA method describes autocorrelation between data. Table 3 shows the methods related to ARIMA and their calculation. A series of random walk (RW) is actually the difference between two consecutive data. The moving average (MA) method uses a regression-like model in the average of past prediction errors instead of using past ones. In the autoregression (AR) method, the prediction is performed using a linear combination of variable past values. ARIMA is composed of a combination of RW, MA and AR methods [1].

Table 2. The calculation of different groups of forecast methods based on a variety of trend and seasonality [1]

Trend	Seasonal N	Seasonal A	Seasonal M
N	$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$ $\hat{y}_{t+h\|t} = \ell_t$	$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = \ell_t + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t/\ell_{t-1}) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = \ell_t s_{t-m+h_m^+}$
A	$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $\hat{y}_{t+h\|t} = \ell_t + hb_t$	$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = \ell_t + hb_t + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + b_{t-1})) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = (\ell_t + hb_t)s_{t-m+h_m^+}$
A_d	$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t$	$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + \phi b_{t-1})) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = (\ell_t + \phi_h b_t)s_{t-m+h_m^+}$
M	$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}$ $\hat{y}_{t+h\|t} = \ell_t b_t^h$	$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = \ell_t b_t^h + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1}b_{t-1})) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = \ell_t b_t^h s_{t-m+h_m^+}$
M_d	$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h}$	$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}^{\phi}) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h} + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $s_t = \gamma(y_t/(\ell_{t-1}b_{t-1}^{\phi})) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h} s_{t-m+h_m^+}$

Table 3. Comparing AR, MR, RW, ARIMA methods

Method	calculation	Function	ARIMA form
RW	$y'_t = y_t - y_{t-1}$	$(1-B)^d y_t$	ARIMA(0,1,0)
MA	$y_t = c + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}$	MA(q)	ARIMA(0,0,q)
AR	$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t$	AR(p)	ARIMA(p ,0,0)
ARIMA	$y'_t = c + \varphi_1 y'_{t-1} + \cdots + \varphi_p y'_{t-p} + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} + \varepsilon_t$	ARIMA(p,d,q)	ARIMA(p,d,q)

Artificial neural networks

Artificial neural networks (ANNs) mimic the function of data processing in the human brain. The human brain is made up of billions of interconnected neurons, each of which is used to store small amounts of information. But by connecting them together, large and complex volumes of information can be stored and processed. To predict time series, an artificial neural network is created that receives a number of historical observations as input to predict the next step. Several runs are performed to update the weights that determine the output result in order to minimize the overall prediction error. It is possible to create a complex nonlinear relationship between input and output in ANNs [19].

Unlike exponential smoothing models, ANNs are not limited to patterns in the data, such as trends or seasonality. It learns these patterns directly from the data and can even model nonlinear data patterns. Interdependence in our complex world is rarely linear, in fact, nonlinearity exists in many real-world forecasts. It can approximate any continuous linear and nonlinear function with any level of precision [20][21].

In the literature, various neural network structures have been presented such as adaptive linear neuron, multi-layer perceptron, and radial basis functions. But most research uses a multilayer perceptron structure to predict demand [22].

To predict a three-layer neural network proposed. It consists of an input, hidden layers, and an output layer structure which is taught from the backpropagation algorithm. During training, the information obtained from the input neurons is multiplied by its assigned weight. The result enters the activation function, which is activated when it reaches a certain threshold, and then the signal is sent to the output neurons. In each round, the prediction error is calculated and returned to the neural network .In using artificial ANNs to predict, first, the structure is determined, and then the training steps and determining the optimal weights is performed .In determining the structure, it is necessary to determine the number of input, output, hidden layers and the relationship between nodes, transfer function, training and testing samples, training algorithm, and performance measures [23]. Kumar and Herbert [24] demonstrate backpropagation training performance has been meaningly better than other training methods. The performance of ANNs can be affected by the number of hidden. Zang [25] determined the sufficient amount of hidden layer to cope with nonlinearity.

Forecast errors

Forecast error measurement, in addition to providing valuable information about the size and variability of random components, is also useful for selecting the appropriate forecasting model and considering the required precautionary reserves. Table 4 presents the common errors in forecasting, how to calculate them, and permitted interval.

Table 4. Forecast Errors [5]

Error Name	Formula	Description
The Root Mean Squared Error (RMSE)	$\text{RMSE}_t = \sqrt{\sum_{i=1}^{t} (E_i - D_i)^2/t}$	The smaller the better
The Mean Absolute Deviation (MAD)	$\text{MAD}_t = (\sum_{i=1}^{t} \lvert E_i \rvert)/t$	The smaller the better
The Mean Absolute Percentage Error (MAPE)	$\text{MAPE}_t = (\sum_{i=1}^{t} [\lvert E_i/D_i \rvert \times 100])/t$	≤ 50
The Mean Percentage Error (MPE)	$\text{MPE}_t = \sum_{i=1}^{t} [E_i/D_i]/t$	The smaller the better
Bias	$\text{bias}_t = \sum_{i=1}^{t} [E_i]$	Smaller numbers are better
Theil's U	Statistics	≤ 1
Ljung-Box Q	Statistics	Applicable in ARIMA model
Tracking Signal	$TS_t = \text{bias} / \text{MAD}_t$	Between -6 to +6

Association Rules

Exploration of association rules is one of the most popular data mining techniques that examines the relationships between groups of items in a particular transaction database. This method has many applications in areas such as market basket analysis, customer profile analysis, recommendation systems, and collaborative filtering [26][27].

Popular models in this field use several iterations to quantify the relationship between items and find the rules of communication For instance if you consider shopping basket analysis, the frequency of occurrence of various items in a set of transactions, is an alternative to measure customer behavior. Finding these rules in the data helps a lot in making important strategic decisions such as the shelf position or the packaging of items that are bought together. There are many algorithms such as Apriori [28], FP growths [29] or Eclat [30] that explore associations based on the number of iterations in the data. Another use for these connections is to find conflicting rules. For example, if a product is purchased, no similar or alternative product will be purchased. Discovering these rules provides valuable information for decision-making. Substitute and complementary goods will play an important role in reducing warehouse costs and creating a marketable portfolio [31].

The first or support equation indicates the number of times A and B have been purchased together. Eq. 28 shows the probability of buying B provided that A has been bought. The third equation shows the positive (lift> 1), negative (lift <1), or neutral (lift = 1) effect of A on selection B. In the last equation, the substitution ratio is defined. The smaller value shows that A and B can be substituted [1].

$$\text{support}\left(\mathrm{A} \Rightarrow B\right) = \text{support}\left(\mathrm{A} \Rightarrow B\right) = \text{support}(\mathrm{A} \cup B) = p(A \cap B) \tag{27}$$

$$\text{confidence}\left(\mathrm{A} \Rightarrow B\right) = p(B|A) = \frac{p(A \cap B)}{p(A)} \tag{28}$$

$$\text{lift}\left(\mathrm{A} \Rightarrow B\right) = \frac{\text{confidence}\left(\mathrm{A} \Rightarrow B\right)}{p(B)} = \frac{p(A \cap B)}{p(A)p(B)} \tag{29}$$

$$\text{Substitute Ratio} = \frac{p(A \cap B)}{\min(p(A), p(B))} \tag{30}$$

In marketing and consumer behavior, a theoretical model links product, person, and goal to affordance perspective in psychology. According to this theory, the perceived benefits of a product are a function of its performance for a consumer in relation to that person's goal. Therefore, perceived benefits reflect the interaction between product performance and the consumer's goal. Therefore, in satisfying the same goal, goods have the ability to replace each other based on perceived advantages [32]. By combining substitute ratio and understanding customer behavior, alternative goods can be identified.

ABC-XYZ analysis

Product classification is very important for adopting appropriate policies for production, sales, procurement, and inventory management of industrial companies. There are several methods for classifying products that can be used depending on the purpose. The main techniques include ABC analysis (ranking items according to annual turnover or number of sales) and XYZ analysis (fluctuation analysis) [33]. The coefficient of variation, the ratio of the standard deviation to the average consumption, is typically used to specify XYZ categories.

Methodology

In this study, a combination of artificial ANNs and time series was used to predict the sales of products and product groups that are commonly used in short-term and long-term planning. The goal is to find the best prediction model for each product and product family considering the MAPE. The steps of the prediction algorithm are as described below:

1. Communication with Oracle Database of organization and preparing a suitable query to extract the essential data

2. Save information in CSV format for transferring to R software
3. Data processing with forecast libraries, nnfor, neuarlnet
    i. Enter sales records of the first product and find the best time series with the least MAPE
    ii. Forecast creation using artificial neural network and MAPE calculating
    iii. Comparison of the best time series error results with the neural network and selection of the best validation model of the selected model with TS index
    iv. Forecast for the next 12 months with about 95% confidence with the selected method
    v. Transfer results to Excel for report creation
4. Repeat steps i to v for all products

To run the model to detect the product family, at first, the following steps (substitute product finding algorithm) are performed and then the results are transferred to the third row of the first algorithm.
1. Gather the perceived advantages of products from the opinion of experts
2. Provide appropriate transactional data for data mining
3. Explore the association rules, support, confidence, and lift for each pair of products using the “arules” package in R
4. Calculate the substitute the ratio for each pair of products
5. Group products based on the same identified advantages
6. Identify substitute products
    I. Identify replaceable products with the same identified advantage
    II. Offer replaceable products with unequal advantage to sales experts for approval
7. Create product groups and aggregate product sales records based on created groups

If the average MAPE of all products after the implementation of the first algorithm is higher than the specified limit, the normalized values will enter the model. This normalized value is the difference between the sales of each product from the lowest sales in the same period divided by the difference between the maximum and minimum sales in the same period. This normalized value is calculated and entered into the algorithm. Based on ABC-XYZ analysis and substitute products, the best sales combination is offered. For example, if two alternative products are one of the AX class and the other of the CZ class, the total sales will be transferred to the AX product. This offer is given to sales experts for review and approval. The priority is to increase sales (conversion of products from class C to B and from B to A and in the next stage to reduce fluctuations from Z to Y and from Y to X).

The offers to the sales department can be summarized as follows.
1. Consolidation of substitute products with the same goal with a better sale pattern according to ABC-XYZ analysis
2. Increasing the sales capacity of products or converting products from C to B and from B to A and reduce sales fluctuations or conversion from Z to Y and from Y to X
3. Elimination of CZ Group products

Case study

Recent research has been carried out on one of the product categories of ZARIF-MOSAVAR Industrial Group. It has been established in the year 1984. First, its activities were focused on carpet and fiber goods production. During the last decades, this group extended its activities in various areas such as textiles, chemicals and polymers, Regal petrochemical. Now about 4000 employees are working in this industrial group. As the main carpet manufacturer in Iran and the middle east has been granted numerous honors including Paris international award in 1998. It produces more than 50 million square meters of carpet products.

4-year historical records of 511 non-woven products of the organization were extracted from the flooring products family to provide a forecast for the next 12 months (Table 5). The optimal forecast model for the total sale and individual products sales have been determined based on the least MAPE. The detected model forecast the next year based on the selected method. Normally non-woven products are divided into 4 categories: plain felt, printed, printed-embroidered, and embroidered.

Table 5. Partial part of historical selling data

Row	Year	2016	2016	2016	2016	2016		2017	2017	2017	2017		2019	2019	2019	2019
	Month	1	2	3	4	5	. . .	9	10	11	12	. . .	9	10	11	12
	total	317,338	1,138,154	1,198,130	943,726	2,081,599	. . .	1,376,606	1,522,353	1,110,852	1,836,933	. . .	1,275,695	1,337,932	1,454,493	1,637,299
1	Alder parquet-brown-5		42,418	37,126	25,890	73,499	. . .	95,204	110,176	73,434	142,380	. . .	44,455	56,119	68,432	76,381
2	Shade-cream-5	22,687	46,637	46,121	50,238	111,898	. . .	66,765	74,098	46,308	85,641	. . .	29,505	41,786	50,163	31,021
3	Sharareh-Chocolate-1	15,631	31,615	47,363	56,429	120,026	. . .	53,739	67,349	44,617	67,789	. . .	29,305	43,868	37,151	32,587
4	Alder-chocolate parquet-5		46,455	40,332	27,224	73,237	. . .	51,797	55,315	43,320	75,210	. . .	33,058	19,383	43,240	33,671
5	Diba-Chocolate-1	11,792	30,066	38,229	15,286	50,810	. . .	37,154	37,666	21,058	33,077	. . .	25,543	34,145	33,377	50,393
6	Alder Parquet-Cream-5		12,975	16,138	11,163	41,017	. . .	40,583	44,588	32,999	64,043	. . .	20,528	28,068	36,091	47,334
7	Aristocratic-brown-5	14,694	27,536	37,599	24,508	59,850	. . .	36,473	45,360	35,891	50,345	. . .	9,949	7,247	15,601	3,903
8	Radin-Nescafe-5	3,240	33,959	31,294	12,920	28,876	. . .	33,033	31,743	27,332	33,740	. . .	14,574	18,616	28,479	18,505
9	Flare-Nescafe-1	2,937	9,115	12,384	14,182	28,441	. . .	21,236	29,295	22,031	33,445	. . .	9,396	19,704	16,058	26,225
10	Simin-Chocolate-3	7,797	32,969	30,380	14,752	27,555	. . .	18,637	17,672	9,986	24,938	. . .	26,729	12,968	24,495	10,058
. .	. .	. .	. .	. .	. .	. .	. . .	. .	. .	. .	. .	. . .	. .	. .	. .	. .
196	Naroon-Messi-5						. . .					. . .	3,140	235	21	
197	Fryal-cream-5						. . .					. . .	496	910	1,165	750
198	Nazanin-Abi-3						. . .	1,235	842	9		. . .	470	1,157	847	1,057
199	Breeze-bronze-3						. . .				1,030	. . .	895	102	49	
200	Pine-chocolate-3						. . .					. . .	2,018	8,019	8,238	11,836
201	Silvana-crimson-1						. . .					. . .	568	1,276	759	1,816
202	Mahan-Red -5	1,203	3,591	2,237	1,366	5,509	. . .	1,615		9	163	. . .				
203	Diba-Bronze-1						. . .			514	2,129	. . .				
204	Fryal-Chocolate-5						. . .					. . .	637	373	238	
205	Silvana-Tusi-1						. . .					. . .	285	541	730	1,801
. .	. .	. .	. .	. .	. .	. .	. . .	. .	. .	. .	. .	. . .	. .	. .	. .	. .
501	Mahan-Chocolate-5						. . .					. . .				
502	Chocolate-3						. . .					. . .				
503	Sharareh-Jasmine-1						. . .					. . .				
504	Nasim-Crimson-3						. . .				13	. . .				
505	Golden-camel -3						. . .					. . .				
506	Enamel-red -1						. . .					. . .			12	
507	Golden-Yasi -3						. . .			11		. . .				
508	Samin-bulb -3						. . .			10		. . .				
509	Egyptian-green-black-1						. . .					. . .				
510	Sultan-Chocolate-3						. . .			9		. . .				
511	Sultan-Brown-3						. . .					. . .				

Results and discussion

The prediction was implemented for cumulative sales and individual products. The results reveal that total sales using ANNs have the lowest MAPE error. Fig. 1 shows the historical records, the results of the forecast and the upper and lower limits in the range of 95% confidence for this forecast. Table 6 presents the error results of different forecasting methods. These methods are ranked from low to high based on MAPE error. If we use the fourth method to forecast total sales, the trend and season components will be as shown in Fig. 2.

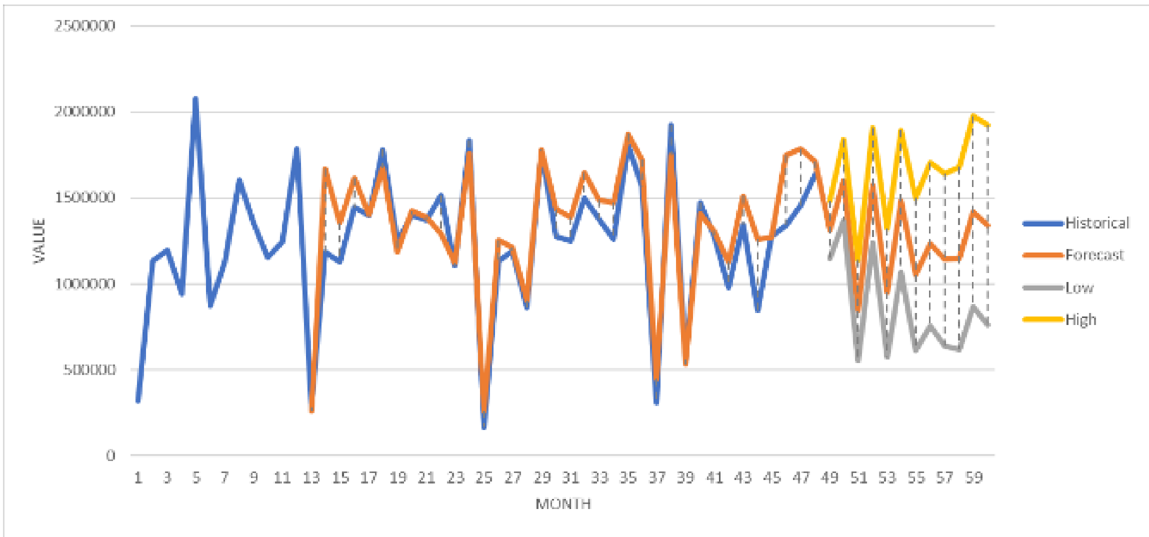

Fig. 1. Results of the best model of total sales forecast with the minimum forecast error

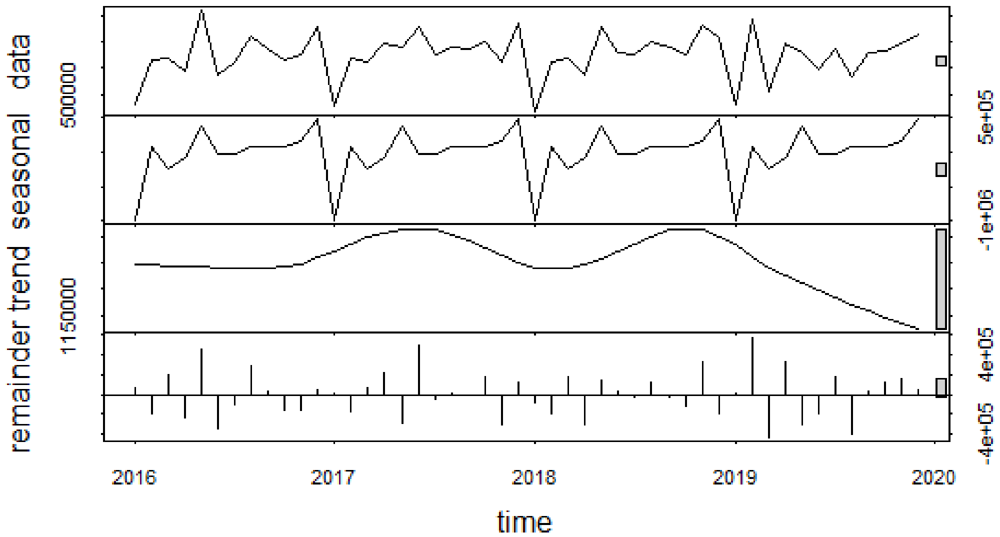

Fig.2. Decomposing the total sale into trend and seasonality components

Table 6. Comparison of forecast errors in different prediction methods

		RMSE	MAD	MAPE	MPE	Theil U
Neural Net	1	176710	130451.2	0.128	0.096	0.235
Multiplicative Seasonal Method	2	351210.7	238453.5	0.217	0.056	0.337
Holt & Winter for Add. Seasonal	3	351210.5	238502.6	0.217	0.056	0.337
Additive Seasonal Method	4	351210.5	238502.6	0.217	0.056	0.337
Damped for Multi. Seasonal	5	351210.5	238502.6	0.217	0.056	0.337
Damped for Add. Seasonal	6	351210.5	238502.6	0.217	0.056	0.337
Holt & Winter for Mult. Seasonal	7	351210.5	238502.6	0.217	0.056	0.337
Linear Exponential Smoothing	8	415270.7	297067.1	0.454	0.245	0.333
Single Exponential Smoothing	9	419065.4	307864.1	0.459	0.227	0.349
Triple Exponential Smoothing	10	422676.5	309105.5	0.469	0.243	0.344

		RMSE	MAD	MAPE	MPE	Theil U
Polynominal Growth	11	408953.6	293937.3	0.487	0.311	0.349
Linear Growth	12	417658.2	296580	0.497	0.321	0.318
Quadratic Growth	13	415872.3	296792.7	0.498	0.322	0.326
Damped Exponential Smoothing	14	434893.4	306619.9	0.5	0.303	0.346
Moving Average	15	510740.1	383698.3	0.6	0.349	0.461
Double Moving Average	16	562804.2	434526.5	0.685	0.389	0.525

Table 7 demonstrates the mean error results for each product in the normalized and not normalized data. 42% reduction in MAPE error was created by normalizing the data. It is worth-noting that outlier data were deleted before entering the prediction algorithm. Normalization at this stage is almost equal to the share of sales of each product in total sales per month.

Table 7. Comparison of average forecast errors between the normalized and not normalized data

	RMSE	MAD	MAPE	MPE	Theil U
Not-Normalized	9195.152	7248.48	0.786727	0.393303	0.572091

	RMSE	MAD	MAPE	MPE	Theil U
Normalized	0.004818	0.003515	0.454909	0.183667	0.685848

For annual and aggregate production planning, the sales of the product families are typically considered. The historical sales records have been aggregated based on 4 common groups and entered into the algorithm. The error of each group is shown in Table 8.

Table 8. Average of prediction errors considering the former categories

Series	Method	RMSE	MAD	MAPE	MPE	Theil U
Printed	Double Moving Average	0.064	0.048	0.124	0.009	0.952
Printed-Brocade	Damped Exponential Smoothing	0.008	0.006	-	-	-
Brocade	Holt & Winter for Mult. Seasonal	0.04	0.033	0.167	0.074	1.212
Simple felt	Holt & Winter for Mult. Seasonal	0.06	0.049	0.14	0.014	0.481

Table 9 demonstrates the results of the implementation of the third step of the algorithm for finding substitute products for each pair of products. The small values of the substitute ratio indicate that the two products can be replaced.

Table 9. Part of the results of exploring association rules in transactional data with substitute ratio for each paired product

Antecedent Product	Consequence Product	Confidence	Support	Lift	p(a)	p(c)	Substitute Ratio
Moscow-cream-5	Moscow-Chocolate-5	0.839	0.022	16.483	0.002	0.003	12.282
Shade-blue -5	Shade-cream-5	0.816	0.097	1.777	0.008	0.031	11.948
Shade-red -5	Shade-cream-5	0.815	0.070	1.775	0.006	0.031	11.932
Olive shade -5	Shade-cream-5	0.808	0.036	1.759	0.003	0.031	11.829

Antecedent Product	Consequence Product	Confidence	Support	Lift	p(a)	p(c)	Substitute Ratio
Shadow	Shade-cream-5	0.803	0.058	1.747	0.005	0.031	11.747
Flare-brown-1	Sharareh-Chocolate-1	0.780	0.132	1.618	0.012	0.033	11.418
Sharareh-Yasi-1	Sharareh-Chocolate-1	0.774	0.043	1.606	0.004	0.033	11.334
Flare-Nescafe-1	Sharareh-Chocolate-1	0.771	0.168	1.598	0.015	0.033	11.279
Sahar-Nescafe-1	Sahar-Chocolate-1	0.770	0.032	7.473	0.003	0.007	11.264
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
Chemistry-chocolate-1	Diba-Chocolate-1	0.247	0.025	0.867	0.007	0.019	3.614
Aristocratic-cream-5	Diba-Brown-1	0.247	0.021	1.278	0.006	0.013	3.612
Nasim-Nescafe-3	Aristocratic-crimson-5	0.247	0.030	1.396	0.008	0.012	3.610
Shade-red -5	Breeze-brown-3	0.246	0.021	1.480	0.006	0.011	3.606
Breeze-brown-3	Radin-Nescafe-5	0.246	0.041	0.954	0.011	0.018	3.605
Maral-brown-1	Radin-Nescafe-5	0.246	0.023	0.953	0.006	0.018	3.604
Dandelion-chocolate-3	Shahrzad-Chocolate-5	0.246	0.028	1.710	0.008	0.010	3.600
Shade-red -5	Aristocratic-crimson-5	0.245	0.021	1.389	0.006	0.012	3.592
Chemistry-chocolate-1	Radin-Nescafe-5	0.245	0.024	0.950	0.007	0.018	3.590
Simin-Nescafe-3	Diba-Nescafe-1	0.245	0.032	1.425	0.009	0.012	3.590
Alder-chocolate parquet-5	Radin-Nescafe-5	0.245	0.096	0.948	0.027	0.018	5.447
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
Alder-chocolate parquet-5	Rasha-brown-5	0.051	0.020	1.085	0.027	0.003	6.236
Alder-chocolate parquet-5	Sharareh-Yasi-1	0.051	0.020	0.909	0.027	0.004	5.225
Sharareh-Chocolate-1	Diba-Yasi-1	0.051	0.025	0.831	0.033	0.004	5.867
Alder parquet-brown-5	Enamel-crimson-1	0.051	0.027	0.798	0.036	0.004	6.232
Sharareh-Chocolate-1	Atlas-Nescafe-5	0.050	0.024	0.739	0.033	0.005	5.216
Sharareh-Chocolate-1	Diba-Tusi-1	0.050	0.024	0.839	0.033	0.004	5.919
Alder parquet-brown-5	Sharareh-Yasi-1	0.050	0.027	0.898	0.036	0.004	7.016
Shade-cream-5	Dandelion-brown-3	0.050	0.023	0.841	0.031	0.004	5.656
Shade-cream-5	Narun-cream-5	0.050	0.023	0.920	0.031	0.004	6.188
Alder parquet-brown-5	Kish-cream-5	0.050	0.027	0.761	0.036	0.005	5.943

After identifying new categories with the same advantages and aggregating their historical records into one group. Similarity perceived by customers are price and specification of product. the prediction algorithm was implemented for these ten new groups. Table 10 displays the prediction error results for these new categories. MAPE error comparison of these new categories with 4 previous groups is shown in Table 11. Increasing the forecast accuracy is more cost-effective for using aggregate planning and will require less safety stock inventory.

Table 10. Average forecast errors in new explored groups

Series	Method	RMSE	MAD	MAPE	MPE	Theil U
S1	Neural Net	0.017	0.012	0.113	0.038	0.57
S12	Neural Net	0.008	0.006	0.216	0.083	0.425
S13	Polynominal Growth	0.032	0.025	0.101	0.016	0.86
S2	Expert Neural Net	0.009	0.007	0.135	0.082	0.803
S21	Polynominal Growth	0.014	0.01	0.114	0.022	0.768
S23	Neural Net	0.009	0.006	0.105	0.006	0.707
S3	Quadratic Growth	0.029	0.022	0.125	0.025	0.759
S31	Polynominal Growth	0.017	0.013	0.108	0.02	0.746
S32	Damped for Add. Seasonal	0.015	0.012	0.244	0.016	0.867
S5	Triple Exponential Smoothing	0.006	0.003	-	-	-

Table 11. Average prediction errors in newly created and previous groups

	RMSE	MAD	MAPE	MPE	Theil U
New Group	**0.0156**	0.0116	0.1261	0.034222	0.722778
	RMSE	**MAD**	**MAPE**	**MPE**	**Theil U**
Previous Group	**0.037333**	0.029	0.1455	0.0415	0.881667

Employing the ABC-XYZ analysis along with the new groups, we have an appropriate guide to select the right products to replace. For instance, Table 12 shows that in the first category (S1), 12% share of sales belongs to the CZ group. The sales share of this group can completely be transferred to BX. In this way, BX can be upgraded to AX as the volume of this group increases. The last group (S5) includes 42% CZ groups. By transferring production and share of this group CX can be improved to BX.

Table 12. Results of ABC-XYZ analysis

Group	AX	AY	BX	BY	BZ	CX	CY	CZ	Wight
S1	0%	7%	18%	32%	0%	0%	30%	12%	13%
S12	23%	0%	53%	12%	0%	0%	12%	0%	3%
S13	42%	39%	14%	4%	0%	0%	1%	0%	27%
S2	0%	0%	19%	43%	1%	10%	25%	2%	6%
S21	0%	0%	47%	53%	0%	0%	0%	0%	9%
S23	0%	0%	25%	75%	0%	0%	0%	0%	6%
S3	28%	36%	15%	11%	0%	2%	7%	0%	19%
S31	54%	28%	19%	0%	0%	0%	0%	0%	12%
S32	56%	27%	17%	0%	0%	0%	0%	0%	5%
S5	0%	0%	0%	0%	0%	58%	0%	42%	1%

Managerial implications

As a direct result of this research, inventory reduction are provided for the company. Confusing historical data can lead to wrong decision solved by using mining rules. Superior products are determined and substituted for others. This can help the organization to decrease demand fluctuation and inevitable set-up times. Moreover, the production capacity is focused on fertile products with higher sales capacity and lower variance.

Conclusion

Sales and demand forecasting contribute greatly to the proper creation of strategic, tactical, and operational decisions in industrial complexes. Data mining techniques such as time series and artificial neural networks have been widely utilized in the demand forecast. In this research, combinatorial use of data mining techniques is applied to cope with multi-product and product family forecasting. First, time-series methods and artificial neural networks have been applied to forecast the sale of all products belonging to the non-woven category of Zarif Mosavar. There are exist a large variety of products with different trends. It is important to choose the appropriate method for each product. To minimize forecasting error, the appropriate method among time series and the artificial neural network was selected to predict the future demand of each product. Product family prediction is also interested in aggregate production planning. The predictions of product families are also produced by the best method. However, replacing the order from one product to another similar product frequently occurs whilst there is no inventory. To cope with this issue, substitute products were identified by exploring the association rules. Then by prioritizing the similar usage, discovered substitute products were integrated and entered into the prediction model. By means of this approach, 10 new groups have been developed. The MAPE of the new groups was reduced compared to the conventional ones. Detection of superior products has been performed by the ABX-XYZ analysis. It means the ABC-XYZ supports finding which substitute product must remain and which of them must be replaced in favor of dominant products. In this way, valuable suggestions have been provided for increasing the sale amount of superior products and decrease demand variations. The organization benefits from the results of this research by decreasing the safety stock inventory and provide more marketable products.

References

[1] Hyndman, R. J., and Athanasopoulos, G. "*Forecasting : Principles and Practice*", (2019).

[2] Gonçalves, J. N. C.; Cortez, P.; Carvalho, M. S.; Frazão, N. M. "A multivariate approach for multi-step demand forecasting in assembly industries: Empirical evidence from an automotive supply chain". Decision Support Systems, 113452, (2020).

[3] del Campo-Ávila, J.; Takilalte, A.; Bifet, A.; Mora-López, L. "Binding data mining and expert knowledge for one-day-ahead prediction of hourly global solar radiation", Expert Systems with Applications, 114147, (2020)

[4] Van Nguyen, T.; Zhou, L.; Chong, A. Y. L.; Li, B.; Pu, X. "Predicting customer demand for remanufactured products: A data-mining approach". European Journal of Operational Research, 281(3), pp.543–558, (2020).

[5] Chopra, S., and Meindl, P. "*Supply chain management: Strategy, planning, and operation* (6th ed.)", Upper Saddle River, New Jersey: Pearson Education, Inc. (2016)

[6] Wheelwright, S.C. and Makridakis, S.G. "Forecasting methods for management(5th ed.)". Wiley, (1989)

[7] Mentzer, JohnT. "Forecasting with adaptive extended exponential smoothing", Journal of the Academy of Marketing Science, 16(3-4), pp.62-70, (1988).

[8] Pantazopoulos, Sotiris N., and Pappis, Costas P. “A new adaptive method for extrapolative forecasting algorithms”, European Journal of Operational Research, 94(1), pp.106-111, (1996).
[9] Roberts, S. D., and R. Reed. "The Development of a Self-Adaptive Forecasting Technique", AIIE Transactions I (No. 4), 314-322, (1969).
[10] Hyndman, R. J.; Koehler, A. B.; Snyder, R. D.; Grose, S. “A state space framework for automatic forecasting using exponential smoothing methods”. International Journal of Forecasting, 18, pp.439–454, (2002).
[11] Taylor, J. W. “Exponential smoothing with a damped multiplicative trend. International Journal of Forecasting”, 19, pp.273– 289, (2003).
[12] Muth, J. F. “Optimal properties of exponentially weighted forecasts”, Journal of the American Statistical Association, 55, pp.299– 306, (1960).
[13] Holt, C. C. “Forecasting seasonals and trends by exponentially weighted averages”, O.N.R. Memorandum 52/1957, Carnegie Institute of Technology. Reprinted with discussion in 2004. International Journal of Forecasting, 20, pp.5 – 13. (1957).
[14] Winters, P. R. “Forecasting sales by exponentially weighted moving averages”, Management Science 6, pp.324–342, (1960).
[15] Anne B.; Koehler Ralph D.; Snyder J.;Keith Ord, “Forecasting models and prediction intervals for the multiplicative Holt–Winters method”, International Journal of Forecasting, 17, pp. 269-286, (2001)
[16] Hamzaçebi, Coşkun. “Improving artificial neural networks’ performance in seasonal time series forecasting”, Information Sciences, 178(23), pp.4550-4559, (2008).
[17] G.E.P. Box, G.M. Jenkins, “*Time Series Analysis Forecasting and Control*”, Holden-Day, San Francisco, (1976).
[18] Zhang, G.; Peter, ; Qi, Min. “Neural network forecasting for seasonal and trend time series”, European Journal of Operational Research, 160(2), pp.501-514, (2005).
[19] Williams, T. M. “Adaptive Holt-Winters forecasting”, Journal of the Operational Research Society, 553-560, (1987).
[20] Zhang, G.; Patuwo, B. E.; Hu, M. Y. “Forecasting with artificial neural networks: The state of the art”, International journal of forecasting, 14(1), pp.35-62,(1998).
[21] Hornik, K.; Stinchcombe, M.; White, H. “Multilayer feedforward networks are universal approximators”. Neural networks, 2(5), pp.359-366,(1989).
[22] Funahashi, K.I. “On the approximate realization of continuous mappings by neural networks. Neural networks”, 2(3), pp.183-192, (1989).
[23] Da Costa Lewis, N. “Neural Networks for Time Series Forecasting with R”. (2017).
[24] Yu, Y.; Choi, T.-M.; Hui, C.-L. “An intelligent fast sales forecasing model for fashion products”, Experts systems with application, 38, pp. 7373-7379, (2011).
[25] Kumar, P.; Herbert, M.; Rao, S. “Demand forecasting Using Artificial Neural Network Based on Different Learning Methods: Comparative Analysis”, International journal for research in applied science and engineering technology, 2(4), pp. 364-374, (2014).
[26] Zhang, X. “Time series analysis and prediction by neural networks”, Optimization Methods and Software, 4(2), pp.151-170, (1998).
[27] Aggarwal, Charu C. “*Data Mining: The Textbook*”. Springer, (2015).
[28] Telikani A.; Gandomi A.; Shahbahrami A. “A survey of evolutionary computation for association rule mining”, Information Sciences, 524, pp.318-352, (2020).
[29] Agrawal, R., and Srikant, R. “Fast algorithms for mining association rules. 20th int. conf. very large data bases, VLDB , 1215, pp. 487-499, (1994).
[30] Han, J., Pei, J., & Yin, Y. “Mining frequent patterns without candidate generation”, ACM, 29(2), pp. 1-12, (2000).
[31] Zaki, M. J.; Parthasarathy, S.; Ogihara, M.; Li, W. “New Algorithms for Fast Discovery of Association Rules”. In KDD , 97, pp. 283-286, (1997).
[32] Savasere, A.; Omiecinski, E.; Navathe, S. “Mining for strong negative associations in a large database of customer transactions”, 14th International Conference on IEEE, pp. 494-502, (1998).
[33] Shocker, A. D.; Bayus, B. L.; Kim, N. “Product complements and substitutes in the real world: The relevance of other products”, Journal of Marketing, 68(1), pp.28-40, (2004).

[34] Scholz-Reiter, B.; Heger, J.; Meinecke, C.; Bergmann, J. “Integration of demand forecasts in ABC-XYZ analysis: Practical investigation at an industrial company”, International Journal of Productivity and Performance Management, *61*(4),pp.445–451,(2012).