Int. j. inf. tecnol. (June 2022) 14(4):1937–1947
https://doi.org/10.1007/s41870-022-00875-3

ORIGINAL RESEARCH

Demand forecasting based machine learning algorithms on customer information: an applied approach

Maryam Zohdi[1] · Majid Rafiee[1] · Vahid Kayvanfar[1] · Amirhossein Salamiraad[1]

Received: 13 September 2021 / Accepted: 16 January 2022 / Published online: 12 February 2022
© The Author(s), under exclusive licence to Bharati Vidyapeeth's Institute of Computer Applications and Management 2022

Abstract Demand forecasting has always been a concern for business owners as one of the main activities in supply chain management. Unlike the past, that forecasting was done with the help of a limited amount of information, today, using advanced technologies and data analytics, forecasting is performed with machine learning algorithms and data-driven methods. Patterns and trends of demand, customer information, preferences, suggestions, and post-consumption feedbacks are some types of data that are used in various demand forecasting efforts. Traditional statistical methods and techniques are biased in demand prediction and are not accurate; so, machine learning algorithms as more popular techniques have been replaced in recent researches in the literature. Until the time of conducting this research, extreme learning machine has not been used for intermittent demand prediction, so the novelty of our research is to adopt this algorithm and also other machine learning algorithms such as K-nearest neighbors, decision tree, gradient boosting, and multi-layer perceptron to examine its accuracy and performance in comparison to other approaches. Finally, it is demonstrated that artificial neural network-based methods outperform the other employed techniques through conducting a comparison among the above-mentioned predictors in terms of mean squared error, mean absolute error, coefficient of determination, and computational time. Furthermore, extreme learning machine is the best or at least among the best predictors. At last, for determining whether the obtained results are statistically significant or not, analysis of variance is conducted and the Kolmogorov–Smirnov technique is adopted to test the normality of outcomes.

Keywords Supply chain management · Demand forecasting · Big data analytics · Machine learning algorithms · Customer information

✉ Vahid Kayvanfar
kayvanfar@sharif.edu

1 Department of Industrial Engineering, Sharif University of Technology, Tehran, Iran

1 Introduction

Supply chain is a system including human resources, material resources, and information that tries to establish an integrated flow between supplier and consumer to create value. Each supply chain includes three parts; first, upstream such as suppliers, second is operations, like all the functions and activities in the companies and the third part is downstream including all the channels that transfer and distribute products and services to the final consumer [1]. Demand forecasting is a vital task in managing downstream and a prerequisite of planning, managing, decision making, developing, and creating coordination in a supply chain. This task has always been a concern for industry and business owners, while only the form of forecasting has changed over time. From the beginning of the 1970s till now, purchase prediction was one of the main issues of market research and customer care, it's also one of the main concerns of supply chain managers. Forecasting refers to a set of activities for estimating and acquiring knowledge about future phenomena by using past and present data. Forecasting is also known as pattern analysis. Forecasting demand, without considering consumer information, is not a new approach and was first carried out by Box and Jenkins nearly 50 years ago [2].

Nowadays, with this unstoppable rate of changes in technology and phenomena such as the industry 4.0, data explosion, and content and data which are generated with high speed and volume in social media, web, and search engines, we are facing with a large amount of data. So, traditional statistical methods do not have enough accuracy and reliability and researches have shown that the results have bias, so data-driven methods have been replaced [3]. Whereas the ability to respond to the needs and requirements of the end-user (consumer) is one of the most important goals in any supply chain, it is not possible to achieve this goal without having a vision about his preferences and estimation of the consumer's potential demand in the future. However, fluctuation of demand as a result of several factors can made this activity convoluted and difficult [4]

Prior researches in the literature have shown that attracting a new customer to a chain is 5–10 times more difficult than keeping existing customers [5]. This is an emphasis on demand forecasting based on consumption trends and patterns. Therefore, forecasting demand in a supply chain is a vital task before deciding about selecting suppliers of raw materials, planning production activities, planning and managing inventory, adopting sales and marketing strategies, choosing supply channels, and other factors related to SCM [6].

Over the recent years, phenomena such as globalization of supply chains, increasing product and service diversity, shortening the life cycle of products and a significant increase in competitive markets, have made the prediction more and more important and complex in a supply chain [7]. In the past, the competitive advantage of organizations, companies, manufacturing and service centers was based on factors such as the workspace, number of employees, production rate, and other factors like this, but nowadays what distinguishes the organizations, is the amount of credible information and knowledge about the overall business route. Today, market share belongs to the organizations that have understood the importance of collecting and documenting data from production, supply, delivery, after-sales services, customers' information and then creating a rich and citable database, try to review and process these data, and finally discover hidden information of them. These actions are performed to recognize the patterns of consumption and demand processes to maximize the efficiency of the supply chain, manage all of the elements and components and finally satisfy the consumer to preserve him in the chain. Here are some acronyms that are used throughout the manuscript, shown in Table 1.

Since the time of conducting this research, ELM has not been adopted in the case of intermittent demand that is occurred periodically, so the novelty of our research is to apply this algorithm to examine its performance in comparison to traditional statistical methods and other common ML algorithms for demand forecasting.

In the next subsection we investigate the relation between BDA and SCM in order to clarify the different aspects of that. Effects and consequences of adopting such methods on the efficiency of supply chain have been also discussed.

1.1 BDA and SCM

Today, with the development of technologies and data explosion, the world is facing a large amount of data. These data include consumers' information, opinions, reviews, feedbacks, suggestions to other consumers, records of previous purchases, etc. The existence of such a large amount of data and a set of tools and approaches have led to the concepts called respectively "Big Data" and "BDA". These data are truly huge with four important attributes, i.e., volume, velocity, veracity, and variety, called 4V. However, in some cases, Value is considered as an additional attribute [8]. Using this type of data is so effective in SCM, due to their high potential for better behavior analysis and better understanding of consumer's needs and demand prediction [7]. Consumer behavior analysis, determining the probability of being loyal or leaving the supply chain by consumer, demand patterns, identifying unmet needs, classifying customers, and so many other information are just some advantages of entering the data analytics into a supply chain and interaction between these two fields. The 3V features have required new and more advanced tools for analyzing and discovering hidden patterns and useful information [9]. Jacobs [10] explains that big data is about data whose size force to look beyond the try-and-error methods. This definition, states that big data cannot be processed using conventional equipment or methods. As online consumer data through search engines like Google and Baidu have become increasingly available and with the fast development of the internet, many researchers have recently started to use big data to improve their forecasts [11]. Schaer et al. [12] have studied existing literature about demand forecasting by means of user-generated data from the internet and analyzing consumer behavior, and also advantages and disadvantages of using this type of data for demand prediction and examining whether the benefits and capacities of these data will be sustained in product's life cycle or not. ML algorithms such as the Adaboost algorithm [13, 14], SVM [14–16], random forest algorithm, NNs [15, 17], have been used in the literature in processing and modeling of these type of data. Other applications of these online data are planning for urban transportation and forecasting demand for intra-urban travels [18], stock price forecasting [17], detecting the influenza epidemic [19], forecasting tourism demand [20], car sales [21], sales forecasting of retailers [22], election results prediction [23, 24], etc.

Table 1 Acronyms and explanations

Acronyms	Explanation	Acronyms	Explanation
SCM	Supply chain management	MLP	Multi-layer perceptron
BDA	Big data analytics	KNN	K-nearest neighbors
ML	Machine learning	RPD	Relative percentage deviation
AI	Artificial intelligence	ANOVA	Analysis of variance
ANN	Artificial neural network	MSE	Mean squared error
DT	Decision tree	MAE	Mean absolute error
ELM	Extreme learning machine	R^2	Coefficient of determination
GB	Gradient boosting		
SVM	Support vector machines		

Applications of these data are not only limited to the prediction of demand but also have been adopted for the prediction of delivery time, resource allocation, and inventory management. In this paper, some of the well-known ML algorithms are adopted for demand forecasting and the results are then reported and compared.

2 Background

As stated, one of the significant features of big data is the large 'volume'. This feature has led to ML algorithms being the alternative for traditional statistical approaches. These algorithms are extremely efficient in processing and modeling these types of data [25]. Generally, ML algorithms divide data into the training set and the test set. The quality of training dataset is the most important factor to reach accurate results. As authors have applied ML tools to forecast Australia's automobile gasoline demand, affirm that, training set selection have significant effect on accuracy [26]. For better processing and modeling results, these algorithms need a large volume of data and big data fulfills this need, as well. In general, according to the existing researches in the literature, these algorithms have dealt well with analyzing and modeling big data. On the other side, using time series analysis techniques has also a bias, caused by demand censoring [7]. This problem, which has been studied in many researches, not only creates adverse effects on demand prediction [27, 28] but can also reduce the accuracy in estimating factors such as price fluctuations, while it has also unfavorable effects on making decisions related to inventory management and control [28]. Different phases of the research are given respectively in Fig. 1 and then related works with different ML algorithms are discussed by introducing and giving some basic definitions about algorithms (Fig. 1).

2.1 DT method

DT is one of the most common methodologies for classification problems. It consists of two main parts including nodes and branches. This algorithm is non-parametric and its ability to process large and complex datasets, handle heavily skewed data, managing missing values and robustness to outliers made it more common and efficient. First, the algorithm divides the dataset into two parts, training set, and validation set. The training dataset and the validation dataset are respectively used for building a DT and choosing the optimal tree size. Variable selection, assessing the relative importance of variables, managing the missing values, prediction, and data manipulation are some of the important applications of DTs. In the last decades have been shown that, random forest which is generated by the combination of multiple DTs, is the best ensemble classifier [29]. The basic components of the structure are: nodes, branches, splitting, stopping, pruning [30].

Demand forecasting in the retail sale has been performed through customer classification by using a DT [31]. This model has been adopted to represent better forecasting of demand to establish a high-performance inventory control system. It helps inventory managers to manage their supply chain performance by reducing delivery days and increasing the service level, simultaneously. Researchers suggested the DT-based classification to analyze the customers' behavior [32]. Presented a model based on discount-oriented promotional offers in retail stores and used a data mining technique to discover the effectiveness of promotion on the purchase volume. The model was established based on a selected customer profile and the purchase behavior of the consumers. The purposed model is a DT, based on purchase classes under the promotion and the results showed that the purchase volume depends on two factors, the revenue and the number of children. So, this research can provide sales managers to determine useful strategies for increasing sales. ML approaches have been adopted for forecasting short-term demand for on-

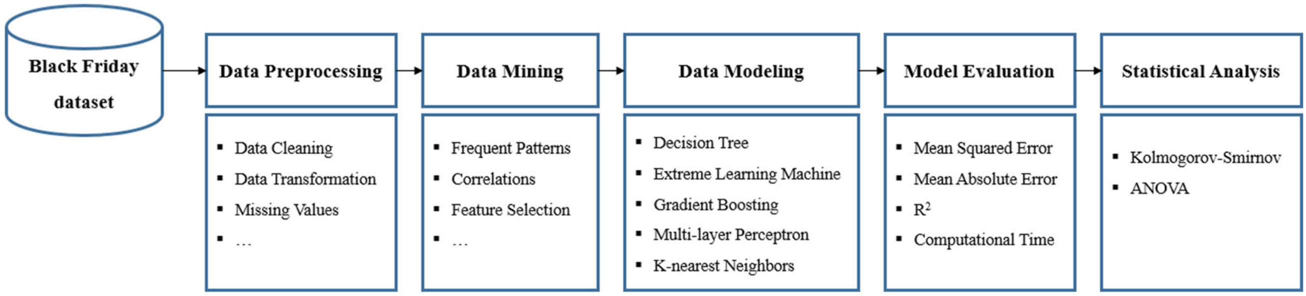


Fig. 1 Phases of the research

demand ride-hailing services [13]. Demand as a function of traffic, price, and weather conditions is the target variable. Single DT, bagged DTs, random forest, boosted DTs, and ANN for regression have been adopted as predictors and then the obtained results compared.

2.2 ELM

An ELM is a novel ANN solution method with a single hidden layer. The most significant characteristic of ELM is initializing input weights and thresholds randomly and then determining output weights by solving the equations. The solution of ELM can be defined as solving the output matrix between the hidden layer and the output layer, to minimize the output error [33]. The learning speed in feed-forward ANNs is so far from what is expected. Low-speed learning is one of the challenges in using ANNs, which can be caused by two factors: first, low speed of gradient-based learning algorithms and second, changing network parameters due to iterating learning process. To solve these issues and increase the speed of learning and faster convergence of the algorithm to find the optimal weight and bias, ELM was used by [34]. Finding local optimal by gradient-based learning algorithms is one of the most important reasons for replacing new learning algorithms. Besides, ELMs have high learning speed, as claimed by [34] are thousand times faster than the old ones, such as backpropagation and also have shown high performance in generalization. As presented in [35] a novel design of interval type-2 fuzzy logic systems by using the theory of ELM for electricity load demand forecasting has been proposed. ELM strategy provides both fast learning of the IT2FLS and optimality of the parameters. Gradient-based algorithms have some disadvantages that the main ones are: (1) slow convergence to the minimum error function by repeating hidden parameters (weights and bias), (2) risk of local minimum points, (3) need to set learning parameters (learning rate and momentum), (4) need to create the number of learning periods and threshold. ELM could overcome these problems. Similarly, authors in [36] adopted exponential smoothing (ES), autoregressive (AR) and autoregressive integrated and moving average (ARIMA) models, multilayer perceptron (MLP) and ELMs to estimate coffee prices. Results based on three measurements showed that the neural networks, especially the ELM, have better performance than the other models. To compare the performance of the backpropagation, ELM and SVM were used by [37], by two datasets considered. For predicting diabetes, SVM and NN have been used with backpropagation and extreme learning. The binary results of modeling on these datasets showed the low performance of the backpropagation algorithm and also stated that the ELM is 1000 times faster than backpropagation and 12 times faster than SVM. In another case study about the prediction of housing in California, it has been shown that ELM is 1000 times faster than the backpropagation and 2000 times faster than the SVM. All three approaches have had the same result on generalization. Unlike gradient descent-based algorithms that only focus on minimizing learning errors, ELM also tries to allocate the smallest weights to the network. Problems like inappropriate learning rates, overfitting, and finding local minimum points, haven't seen in this algorithm. Unlike gradient-based algorithms that use only for derivative activation functions, ELM applies to single-hidden layer feed-forward ANNs.

2.3 GB method

The function consists of a random output or response variable "y", and a set of random input variables $X = \{x_1 \ldots x_n\}$. With considering a training sample, the goal is to find a function, F(x) that maps x to y, such that the loss function is minimized [38]. GB is known as a powerful ML algorithm and a greedy additive strategy. It is an iterative procedure in which at each step we fit the residuals of the previous step using a Bass learning model. GB has received much attention in recent researches and has been adopted in some cases of prediction, such as short term electricity forecast [39], disease forecast [40], predicting short-term subway ridership and prioritizing its influential factors using GBDTs [41], estimation of the price of crude oil [42], short-term power demand prediction using stochastic GB [43].

2.4 MLP method

An MLP is a feed-forward ANN. MLPs could have different architectures based on the expected output and accuracy, but each MLP consists of at least three layers, i.e., input layer, hidden layer(s), and output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP is trained by a supervised learning method called backpropagation. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. An MLP, a radial basis function, and an Elman network were adopted for tourism demand prediction [44]. Then the results have shown that MLP and RBF have better performance and could outperform the Elman network. Monthly water demand prediction using wavelet transform, first-order differencing and linear de-trending techniques based on multilayer perceptron models [45], short-term railway passenger demand forecasting [46], financial time series prediction [47], demand forecasting based on a Moroccan Supermarket dataset [48] are some researches which have been done using MLP.

2.5 KNN method

In pattern recognition, the KNN algorithm is a non-parametric method used for classification and regression [49]. KNN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computations are deferred until classification. A KNN-based model was proposed for predicting the next day's load considering temperature as an important and significant input variable by [50]. The proposed model was tested on real data, and showed high accurate results. A hybrid model, Wavelet denoised-ELM optimized by K-nearest neighbor regression (EWKM), which combines KNN and ELM based on a wavelet de-noising technique was proposed for short-term load forecasting [51]. The proposed model could create a significant improvement in prediction accuracy. Demand forecasting in SCM due to recognizing the pattern and forecasting sporadic demand [52], low-voltage power demand forecasting [53] are some cases that adopted KNN.

3 Related works

Demand forecasting approaches have been used in researches, can be assorted into five categories: deterministic, non-deterministic, statistical methods, AI-based approaches, and the last one is knowledge-based expert systems. Deterministic models include curve fitting, data extrapolation, and smoothing techniques. Kalman filtering, autoregressive moving averages and time-series models are categorized in non-linear models. Also, knowledge-based systems have already shown acceptable performance in predicting demand. But there is something controversial and that is entering new technologies into the supply chain and changing management paradigms from focusing on the initiative into data-driven decision-making. This has led to investigating BDA instead of traditional statistical methods. Traditional statistical methods were used to forecast, do not have enough efficiency, especially when there are large components in the downstream, and variant patterns and behaviors of consumers in the supply chain are observed, especially when these patterns are not linear [54]. In general, implementing traditional methods in forecasting demand is accompanied by simplification and adopting some extra assumptions about the problem, while new methods and tools, extract knowledge by using real-time data, without any changes on the existing hypothesis in the problem or considering new assumptions [55]. Discussed the applicability of advanced ML techniques, including NNs, recurrent NNs, and SVMs, to forecast distorted demand at the end of a supply chain (bullwhip effect). They then compared these approaches with traditional ones, including naive forecasting, moving average, and linear regression. Deep learning (DL), support vector machine (SVM), and artificial neural network (ANN) were used to forecast the transportation-based-CO_2 emission and energy demand in Turkey. Then for comparison, methods were evaluates by R2, RMSE, MAPE, MBE, rRMSE, and MABE. For all ML algorithms, measures reported high accuracy [56]. Demand forecasting has been done to optimize human resource allocation and improve customer service in a medical clinic [57]. In this study, the prediction has been done by combining two methods, feature selecting and deep learning algorithms, which use modified genetic algorithm (GA) to select a feature from a deep NN to predict the demand. To evaluate the validity and accuracy of the model, it has been implemented in a clinic and compared with other predictive methods of demand.

Considering the aforementioned advantages of methods based on data analytics in forecasting in comparison to traditional statistical methods, in the next subsection we will provide some evidences.

3.1 Disadvantage of time series

Murray et al. [58] have stated that time-series models and traditional statistical approaches have low efficiency and accuracy, so data mining methods and clustering have been adopted for forecasting demand. Consumers have been classified according to similar behavior and demand patterns. This has been done to reduce dimensions of the problem and forecast just for each cluster instead of each consumer. Because of the unpredictable and uncertain nature of demand in supply chains, existing forecasting methods are inefficient in discovering irregular patterns and trends of demand. Achieving high accuracy of prediction is the most important goal in all researches done on prediction. According to [59], the prediction accuracy does not only depend on the model, but it also depends on the learning algorithms and feature selection. Feature selection is one of the pre-processing steps in ML algorithms, in which, by deleting the correlated data, one can just keep the least amount of data with the most information, which can be used to reduce the time of processing and implementation of the model, better understanding of data and less required volume for data storage [60]. As stated, the learning method is as effective as building the forecasting model, so it's important to choose an optimal learning algorithm with fast speed and high accuracy.

4 Results

In this study, a dataset from Kaggle website,[1] about customers' personal information including (user ID, product ID, gender, age, occupation, city category, and years of stay in current city, marital status, product category, and purchase) and their demand volume on Black Friday in a retail store were used. As the quality of input data has a high impact on the accuracy of the output, so it's important to remove the inefficiencies of data before modeling. In this research, preprocessing was performed to manage missing data, outliers, inconsistencies, and also extracting patterns and the correlation between attributes to choose the most important and impactful attributes. For having a better insight into our customers and their situation, feature selection was conducted by a data reduction algorithm called the chi-squared test, which is a statistical method that chooses a set of observations highly correlated with the independent variable. After that, we plotted the distribution of customers based on their important attributes. The distribution of customers based on gender (Fig. 2), age (Fig. 3), occupation (Fig. 4), and marital status (Fig. 5) are plotted. Also, the different groups of purchase volumes have been represented in Fig. 6.

As it is obvious, the third interval has been more frequent. The results of analyzing show that the number of male customers and the volume of purchases per male customer are more than female customers. Also, the age group of 26–35 has the highest demand, and there is a positive correlation between purchase volume and being single. According to these results, one can say that marketing and advertising activities could be planned with more focus on these customers.

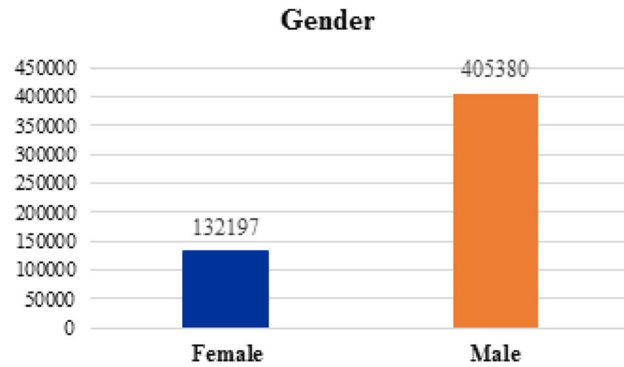


Fig. 2 Distribution of customers based on gender

4.1 Modeling results and evaluation

All instances are implemented in Python 3.6.2 and also Anaconda 3 has been used in this research. All the learning and modeling processes are run on a PC with a 3.5 GHz Intel® Core™ i7-4770 processor and 16 GB RAM. After modeling on dataset, for evaluating the performance of each algorithm, three measures, i.e., MSE, MAE, and R^2 were used. Measures are presented respectively.

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n} = \frac{\sum_{i=1}^{n} e_i^2}{n} \quad (1)$$

$$MAE = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n} \quad (2)$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (3)$$

Table 2 shows the results of five different ML algorithms, where the first column indicates the MSE. As could be seen in Table 2, all of the algorithms yield almost the same results, however based on the exact results which have represented by MSE, MLP and ELM have shown better performance than others in prediction. In Fig. 7, based on MAE, these two algorithms have shown better results again. According to the results of third column, the other measure is R^2 which is a statistical module that provides a measure of how well, the model has been able to replicate observed outcomes, based on the proportion of total variation of outcomes and it ranges over interval 0 and 1. The average value of R^2 is reported in Table 2. It should be pointed out that each algorithm has run five times and the outcomes are presented in Table 2 are the average values. The outcomes are shown in Fig. 8.

According to the obtained results, MLP, ELM and GB have the biggest R^2 values. Since the ANN -based method

[1] http://www.kaggle.com.

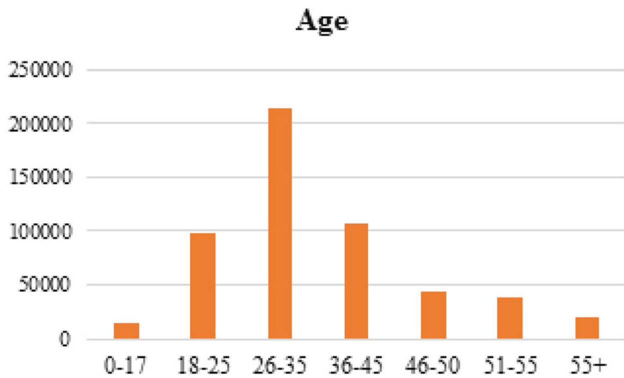


Fig. 3 Distribution of customers based on age

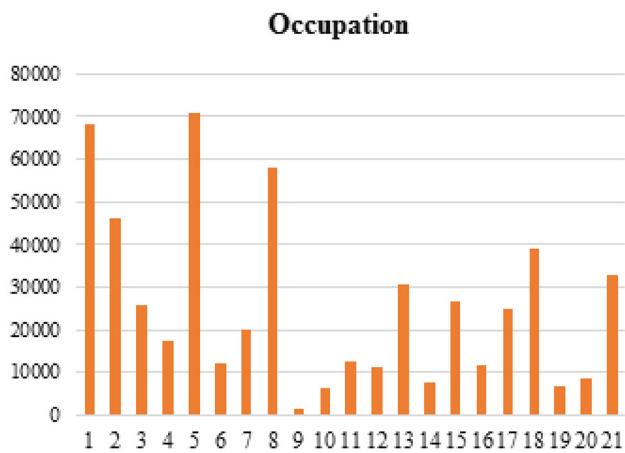
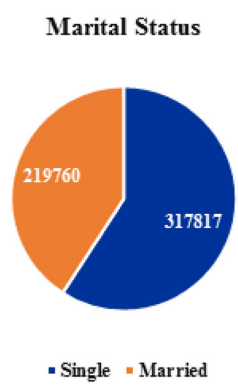


Fig. 4 Distribution of customers based on occupation

Fig. 5 Distribution of customers based on marital status



has a high ability in generalization and managing huge datasets, they improve the modeling procedure more efficiently and can reach more accurate results in forecasting within reasonable computational time (Figs. 7, 8).

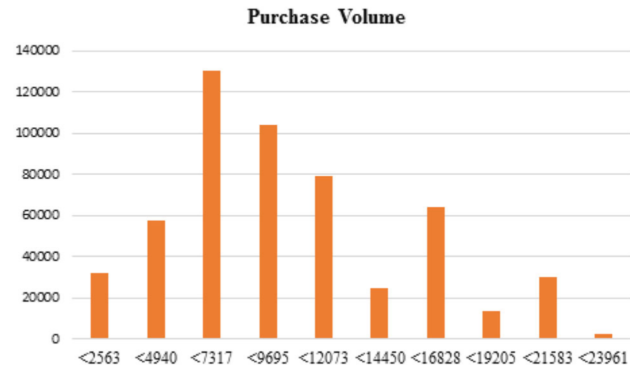


Fig. 6 Distribution of customers based on purchase volume

Table 2 Results of statistical measures

Algorithms	MSE	MAE	R^2
KNN	11091348.65	2455.95	0.4992
GB	9255652.28	2324.94	0.6263
DT	12395358.81	2439.18	0.4877
MLP	9087792.96	2282.62	0.6311
ELM	9088730.24	2198.87	0.6365

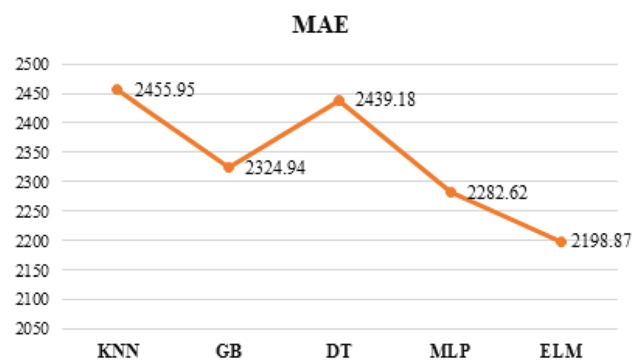


Fig. 7 Comparison of algorithms based on MAE

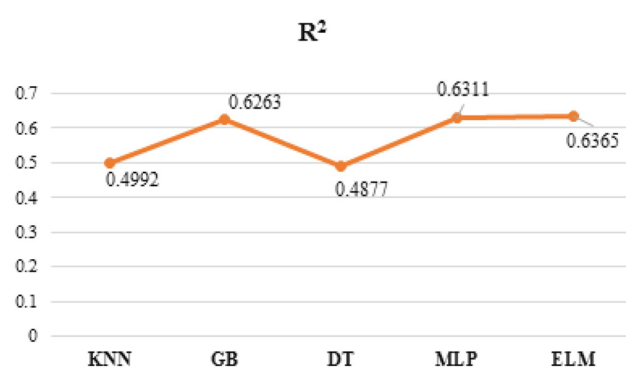


Fig. 8 Comparison of algorithms based on MSE

5 Statistical analysis

To compare the obtained results of the five proposed algorithms, as Eq. (4), the RPD is used as the performance measure and applied over the average results. In Eq. (4), Alg_{mea} signifies the obtained value by each algorithm and Min_{mea} is the min value in each column obtained by running each algorithm. Table 3 shows the relative calculated values. The lower the RPD values, the better the performance of the algorithm.

$$RPD = \frac{Alg_{mea} - Min_{mea}}{Min_{mea}} \tag{4}$$

For determining if the obtained results reported in Table 3 are statistically significant, ANOVA was conducted. Moreover, the Kolmogorov–Smirnov technique is adopted to test the normality of outcomes, as the prerequisite for conducting parametric statistical methods like ANOVA.

Since the obtained P value from Table 4 is greater than the significance level (0.05), therefore the null hypothesis is accepted and accordingly the data are normally distributed. Now, one can use ANOVA, as parametric test. The outcomes of ANOVA are shown in Table 5.

According to the obtained results, there isn't any significant difference between the five proposed algorithms. If there was a statistically significant difference, it would be necessary to use the Tukey test to determine which algorithms are significantly different from each other. Meanwhile, by using the calculated MSE, MAE and R^2, two algorithms showed better performance which are "MLP" and "ELM". This could be evidence illustrating that ANN-based models have acceptable performance in managing and modeling such datasets with huge volume and various attributes. To be more specific, ELM due to its ability and adaptability with large datasets showed higher performance both in providing more accuracy and speed.

Figure 9 shows the processing time of all proposed algorithms to seconds. It should be noted that the data have become dimensionless with the help of Eq. (4). As depicted in Fig. 9, GB, DT, ELM, MLP, and KNN have the highest processing speed, respectively. The KNN algorithm has the highest processing time among the other algorithms, which is not unexpected, by considering its problem-solving approach. It could also be concluded that ELM algorithm falls into the category of fast algorithms for solving these kinds of problems. Therefore, given its speed and accuracy, it can be used as a fast algorithm to solve data-driven problems, especially those dealing with big data like this research.

Table 3 RPD values

Algorithms	MSE (RPD)	MAE (RPD)	R^2 (RPD)
KNN	0.2204	0.1169	0.0236
GB	0.0184	0.0573	0.2842
DT	0.3639	0.1092	0
MLP	0	0.0380	0.2941
ELM	0.0001	0	0.3051

Table 4 Results of one-sample Kolmogorov–Smirnov test

	Value_of_metric
N	15
Normal parameters[a,b]	
Mean	0.122124
Std. deviation	0.1333782
Most extreme differences	
Absolute	0.220
Positive	0.220
Negative	− 0.180
Kolmogorov–Smirnov Z	0.851
Asymp. Sig. (2-tailed)	0.464

[a]Test distribution is normal

[b]Calculated from data

6 Managerial insights

In the last few years, supply chains have been immensely affected with the advent of advanced technologies such as robotics, BDA, AI, block chains, quantum processing, internet of things (IoT), etc. [61]. The most important feature of these technologies is "data explosion", which has led to major changes in various industries and businesses. The supply chains of products and services that are directly in contact with the customer as an end-user, have also been strongly affected by the recent advances in technology. Therefore, supply chain managers must adapt to existing conditions and exploit their organizations' potentials. There are some tips which could led to have a data-driven management;

1) Creating a rich database of supply chain information, including production information, suppliers' information, customers' behaviors, preferences, purchase records, and consumption patterns.
2) Moreover, they must also utilize advanced techniques such as data analysis and AI-based forecasting models and newly-developed tools such as digital marketing and smart customer relationship systems

Table 5 Results of ANOVA test

Value_of_metric					
	Sum of squares	df	Mean square	F	Sig
Between groups	0.005	4	0.001	0.056	0.993
Within groups	0.244	10	0.024		
Total	0.249	14			

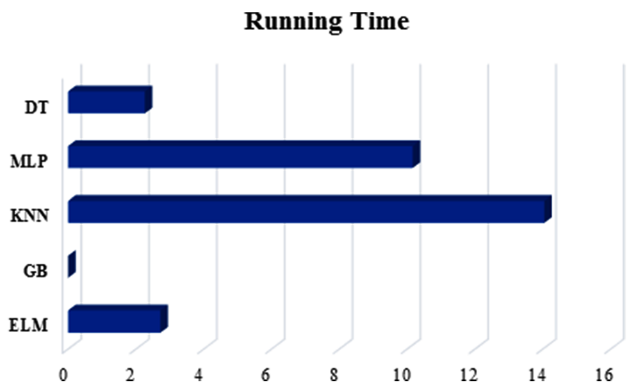


Fig. 9 The running time for five proposed algorithms

to predict potential demand before being expressed by the customers.

3) Needless to say, managers have to give a high priority to the issue of hiring and training data analysts for extracting the most useful information out of the supply chain's raw data to have an integrated data-driven decision-making system.

4) Also replacing new AI-based prediction methods with traditional statistical techniques can improve the quality of prediction. Using ML algorithms, especially ANN-based algorithms that well process large-scale datasets is highly recommended.

7 Conclusions and future studies

Demand forecasting is one of the most vital tasks in managing the supply chain and also a prerequisite of meeting customers' needs. This task has always been a concern for industry and business owners, while only the form of forecasting has changed over time. With the advent of technologies and the expansion of online businesses and online shopping, SCM strategies have changed and so on, the demand forecasting approaches have been revised. Due to these mentioned factors, patterns and trends of demand, customer information, comments and suggestions, post-consumption feedbacks are some types of data that are used in new-developed approaches.

In this research, five ML algorithms; ELM, GB, KNN, MLP and DT, were applied for forecasting demand in a retail store, based on customer information on Black Friday. Since, it is so important for retail stores, especially those with perishable products to predict the number of customers and the volume of their purchases. For example, by predicting the demand volume on Black Friday, a retail store could be able to decide about its surplus inventory all over the year.

Based on the obtained results, MLP, ELM, GB, KNN and DT were the best algorithms in terms of MSE, respectively, while the best performance in terms of MAE was related to ELM, MLP, GB, DT, and KNN. Also, ELM had the higher R^2 value which was 0.6365 and the less value was related to DT (0.4877). Moreover, as the running time could be one of the important criteria for evaluating the performance of algorithms, GB, DT, ELM, MLP and KNN were respectively the fastest in processing the dataset. Taking all results into account mentioned above, for determining the best algorithm, statistical analyses were conducted over the outcomes and the consequences showed that there is no significant statistical difference between the outcomes of the employed algorithms.

For future researches, it is recommended to hybridize different algorithms instead of single-algorithmic models. In this context, with respect to the acceptable performance of the proposed algorithms in data processing and analysis, using hybrid models which are made by combining several algorithms could be a stream for developing single-algorithmic models. Using hybrid models, in addition to eliminating the weaknesses of algorithms, can provide a reinforced model. Also, using the cross-validation method for evaluating the accuracy and validation of the model could be a suggestion to derive a more accurate estimate of model prediction performance.

References

1. Sabbaghi A, Sabbaghi N (2004) Global supply-chain strategy and global competitiveness. Int Bus Econ Res J IBER 3(7). https://doi.org/10.19030/iber.v3i7.3706
2. Box GE et al (2019) Some recent advances in forecasting and control. J R Stat Soc 17:91–109
3. Khan N et al (2014) Big data: survey, technologies, opportunities, and challenges. Sci World J, 712826. https://doi.org/10.1155/2014/712826

4. Kück M, Freitag M (2021) Forecasting of customer demands for production planning by local k-nearest neighbor models. Int J Prod Econ 231, 107837. https://doi.org/10.1016/j.ijpe.2020.107837
5. Martínez A et al (2020) A machine learning framework for customer purchase prediction in the non-contractual setting. Eur J Oper Res 281(3):588–596
6. Jüttner U et al (2007) Demand chain management-integrating marketing and supply chain management. Ind Mark Manag 36(3):377–392
7. Boone T et al (2019) Forecasting sales in the supply chain: consumer analytics in the big data era. Int J Forecast 35(1):170–180
8. Elshawi R et al (2018) Big data systems meet machine learning challenges: towards big data science as a service. Big Data Res 14:1–11
9. Du RY et al (2015) Leveraging trends in online searches for product features in market response modeling. J Mark 79(1):29–43
10. Jacobs A (2009) The pathologies of big data. Commun ACM 52:36–44
11. Silva ES et al (2019) Forecasting tourism demand with denoised neural networks. Ann Tour Res 74:134–154
12. Schaer O et al (2019) Demand forecasting with user-generated online information. Int J Forecast 35(1):197–212
13. Saadi I et al (2017) An investigation into machine learning approaches for forecasting spatio-temporal demand in ride-hailing service, arXiv preprint arXiv:1703.02433
14. Santillana M et al (2015) Combining search, social media, and traditional data sources to improve influenza surveillance. PLoS Comput Biol 11(10):e1004513. https://doi.org/10.1371/journal.pcbi.1004513
15. Yu L et al (2019) Online big data-driven oil consumption forecasting with Google trends. Int J Forecast 35(1):213–223
16. Schneider MJ, Gupta S (2016) Forecasting sales of new and existing products using consumer reviews: a random projections approach. Int J Forecast 32(2):243–256
17. Bollen J et al (2011) Twitter mood predicts the stock market. J Comput Sci 2(1):1–8
18. Zhao Y et al (2018) Improving the approaches of traffic demand forecasting in the big data era. Cities 82:19–26
19. Ginsberg J et al (2009) Detecting influenza epidemics using search engine query data. Nature 457:1012–1014. https://doi.org/10.1038/nature07634
20. Hand C, Judge G (2012) Searching for the picture: forecasting UK cinema admissions using Google Trends data. Appl Econ Lett 19(11):1051–1055. https://doi.org/10.1080/13504851.2011.613744
21. Fantazzini D, Toktamysova Z (2015) Forecasting German car sales using Google data and multivariate models. Int J Prod Econ 170:97–135
22. Abdullah E et al (2017) Papr reduction using scs-slm technique in stfbc mimo-ofdm. J Eng Appl Sci 3218–3221
23. Huberty M (2015) Can we vote with our tweet? On the perennial difficulty of election forecasting with social media. Int J Forecast 31(3):992–1007
24. Mavragani A, Tsagarakis KP (2016) YES or NO: predicting the 2015 GReferendum results using Google Trends. Technol Forecast Soc Chang 109:1–5
25. Al-Jarrah OY et al (2015) Efficient machine learning for big data: a review. Big Data Res 2(3):87–93
26. Li Z et al (2022) Forecasting automobile gasoline demand in Australia using machine learning-based regression. Energy 239 (Part D), 122312. https://doi.org/10.1016/j.energy.2021.122312
27. Wecker WE (1978) Predicting demand from sales data in the presence of stockouts. Manag Sci 24:977–1094
28. Tan B, Karabati S (2004) Can the desired service level be achieved when the demand and lost sales are unobserved?. IIE Trans 345–358
29. Qiu X et al (2017) Oblique random forest ensemble via least square estimation for time series forecasting. Inf Sci 420:249–262
30. Song YY, Ying LU (2015) Decision tree methods: applications for classification and prediction. Shanghai Arch Psychiatry 27(2):130–135
31. Bala PK (2010) Decision tree based demand forecasts for improving inventory performance. In: 2010 IEEE International Conference on Industrial Engineering and Engineering Management, 2010, pp 1926-1930. https://doi.org/10.1109/IEEM.2010.5674628
32. Bala PK (2009) A data mining model for investigating the impact of promotion in retailing. In: 2009 IEEE International Advance Computing Conference, 2009, pp 670-674. https://doi.org/10.1109/IADCC.2009.4809092
33. Liu H et al (2019) Deterministic wind energy forecasting: a review of intelligent predictors and auxiliary methods. Energy Convers Manag 195:328–345
34. Huang G-B et al (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1–3):489–501
35. Hassan S et al (2016) A systematic design of interval type-2 fuzzy logic system using extreme learning machine for electricity load demand forecasting. Electr Power Energy Syst 82:1–10. https://doi.org/10.1016/j.ijepes.2016.03.001
36. Deina C et al (2021) A methodology for coffee price forecasting based on extreme learning machines. Inf Process Agric. https://doi.org/10.1016/j.inpa.2021.07.003
37. Huang G-B, Zhu Q-Y, Siew C-K (2004) Extreme learning machine: a new learning scheme of feedforward neural networks. In: 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541), 2004, vol 2, pp 985–990. https://doi.org/10.1109/IJCNN.2004.1380068
38. Friedman JH (2002) Stochastic gradient boosting. Comput Stat Data Anal 38(4):367–378
39. Mayrink V, Hippert HS (2015) A hybrid method using exponential smoothing and gradient boosting for electrical short-term load forecasting. In: 2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI), 2016, pp 1-6. https://doi.org/10.1109/LA-CCI.2016.7885697
40. Chen X et al (2018) EGBMMDA: extreme gradient boosting machine for MiRNA-disease association prediction. Cell Death Dis 9,3. https://doi.org/10.1038/s41419-017-0003-x
41. Ding C et al (2016) Predicting short-term subway ridership and prioritizing its influential factors using gradient boosting decision trees. Sustainability 8(11):1100. https://doi.org/10.3390/su8111100
42. Gumus M, Kiran MS (2017) Crude oil price forecasting using XGBoost. In: 2017 International Conference on Computer Science and Engineering (UBMK), 2017, pp 1100–1103. https://doi.org/10.1109/UBMK.2017.8093500
43. Nassif AB (2016) Short term power demand prediction using stochastic gradient boosting. In: 2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA), 2016, pp 1–4. https://doi.org/10.1109/ICEDSA.2016.7818510
44. Claveria O et al (2014) Tourism demand forecasting with neural network models: different ways of treating information. Int J Tour Res 17(5):492–500
45. Altunkaynak A et al (2018) Monthly water demand prediction using wavelet transform, first-order differencing and linear detrending techniques based on multilayer perceptron models. Urban Water J 15(2):177–181. https://doi.org/10.1080/1573062X.2018.1424219

46. Tsai T-H et al (2009) Neural network based temporal feature models for short-term railway passenger demand forecasting. Expert Syst Appl 36(2):3728–3736
47. Ravi V et al (2017) Financial time series prediction using hybrids of chaos theory, multi-layer perceptron and multi-objective evolutionary algorithms. Swarm Evol Comput 36:136–149
48. Slimani I, Farissi IE, Achchab S (2015) Artificial neural networks for demand forecasting: application using Moroccan supermarket data. In: 2015 15th International Conference on Intelligent Systems Design and Applications (ISDA), 2015, pp 266–271. https://doi.org/10.1109/ISDA.2015.7489236
49. Altman NS (1992) An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat 6(3):175–185
50. Zhang et al (2016) A composite k-nearest neighbor model for day-ahead load forecasting with limited temperature forecasts. In: 2016 IEEE Power and Energy Society General Meeting (PESGM), 2016, pp 1-5. https://doi.org/10.1109/PESGM.2016.7741097
51. Li W et al (2017) A novel hybrid model based on extreme learning machine, k-nearest neighbor regression and wavelet denoising applied to short-term electric load forecasting. Energies 10(5):694. https://doi.org/10.3390/en10050694
52. Nikolopoulos KI et al (2016) Forecasting supply chain sporadic demand with nearest neighbor approaches. Int J Prod Econ 177:139–148
53. Valgaev O, Kupzog F, Schmeck H (2016) Low-voltage power demand forecasting using K-nearest neighbors approach. In: 2016 IEEE Innovative Smart Grid Technologies - Asia (ISGT-Asia). http://dx.doi.org/10.1109/ISGT-Asia.2016.7796525
54. Du RY et al (2015) Leveraging trends in online searches for product features in market response modeling. J Mark 79(1):29–43. https://doi.org/10.1509/jm.12.0459
55. Carbonneau R et al (2008) Application of machine learning techniques for supply chain demand forecasting. Eur J Oper Res 184(3):1140–1154
56. Ağbulut Ü (2022) Forecasting of transportation-related energy demand and CO_2 emissions in Turkey with different machine learning algorithms. Sustain Prod Consum 29:141–157
57. Jiang S et al (2017) Modified genetic algorithm-based feature selection combined with pre-trained deep neural network for demand forecasting in outpatient department. Expert Syst Appl 82:216–230
58. Murray PW et al (2015) Forecasting supply chain demand by clustering customers. IFAC Pap OnLine 48(3):1834–1839
59. Chandrashekar G, Sahin F (2014) A survey on feature selection methods. Comput Electr Eng 40(1):16–28
60. Ghasab MAJ et al (2015) Feature decision-making ant colony optimization system for an automated recognition of plant species. Expert Syst Appl 42(5):2361–2370
61. Jin X et al (2015) Significance and challenges of big data research. Big Data Res 2(2):59–64