Machine Learning Fusion and Data Analytics Models for Demand Forecasting in the Automotive Industry: A Comparative Study

Esraa Kamal[*1], Amal F. Abdel-Gawad[1], Basem Ibraheem[2], Shereen Zaki[1]
[1]Decision Support Department, Faculty of Computers and Informatics Zagazig University, Zagazig, 44519, Egypt
[2]Computer and Systems Department, Electronics Research Institute ,Giza , Egypt
Emails: esraakamal183@gmail.com; amgawad2001@yahoo.com; Basem@eri.sci.eg; szsoliman@zu.edu.eg

*Correspondence: esraakamal183@gmail.com

Abstract

Demand forecasting is a crucial aspect of managing the supply chain, as it helps companies optimize inventory levels and minimize expenses related to inventory shortages. In recent years, machine learning (ML) algorithms have gained popularity for demand forecasting, as they can handle large and complex datasets and provide accurate predictions. Precise demand prediction for car brands is vital for companies to minimize costs and prevent inventory shortages. The demand for distributing cars is a critical component of inventory management. However, estimating demand for new car sales is difficult due to its continuous nature. To address this challenge, a study was conducted to train, test, and compare the performance of five machine learning algorithms (Random Forest, Multiple Linear Regression, k-Nearest Neighbors, Extreme Gradient Boosting, and Support Vector Machine) using a benchmark dataset. Among all the experiments, the Support Vector Machine algorithm achieved the highest accuracy score of 71.42%. Moreover, Multiple Linear Regression performed well, with an accuracy score of 66.66%. On the other hand, the Extreme Gradient Boosting algorithm had the lowest accuracy score of 42.85%. All experiments used a train-test split of 75/25.

Keywords: Demand Forecasting; Machine Learning; Multiple Linear Regression; Support Vector Machine; K-nearest Neighbors; Random Forest; Extreme Gradient Boosting.

1. Introduction

Forecasting demand is a crucial strategy for managing uncertainty in the supply chain [1,2] and is considered an integral part of supply chain management (SCM). Effective demand forecasting can provide numerous short- and long-term benefits to the organization, such as improving customer satisfaction, reducing inventory stockouts, enhancing production planning, minimizing the need for safety stock, improving transportation management, optimizing pricing and promotion management, negotiating better terms with suppliers, and establishing sales strategies [3,4]. Demand forecasting involves estimating the goods and services that customers will purchase in the near future. Accurate demand forecasting with minimal error margins facilitates important SCM operations such as budgeting, financial planning, sales and marketing plans, raw material supply, production planning, and risk assessment planning [5,6]. Demand forecasting can benefit both suppliers and customers in terms of purchasing raw materials and managing inventory levels. By organizing raw material and product inventory levels, businesses can create a production plan based on customer orders, particularly for medium- and large-sized companies with extensive distribution networks [7,4]. Improving demand forecasting

Doi: https://doi.org/10.54216/FPA.120102
Received: January 09, 2023 Revised: April 08, 2023 Accepted: June 03, 2023

can also stabilize inventories in the network's logistics and enhance transport services. To manage product supply chains effectively, there is a growing focus on analyzing consumer behavior and preferences by utilizing forecasts derived from customer data and transaction records [8,9]. Undoubtedly, one of the biggest and most important sectors in the world is the automotive industry. It involves a wide range of businesses engaged in the planning, creation, production, sale, and marketing of automobiles and their replacement spare parts[10,11]. Globally, the automobile industry is considered to be a pioneer in manufacturing and supply chain management[12]. Using machine learning(ML) and artificial intelligence (AI) to forecast demand has received a lot of attention recently. When compared to conventional techniques of demand forecasting, ML-based forecasting powered by predictive analytics has assisted businesses in improving consumer interaction and producing more precise demand projections when they grow into new markets or channels[13]. The use of ML enables us to change the model's domain of application. It is possible to fit models to larger datasets by combining datasets produced from time series, which increases the likelihood that noise will be isolated from real demand patterns while lowering the danger of overfitting[14,15]. This study offers some helpful insights into predicting car brand demand. For this purpose, the used dataset was gathered from Kaggle then five different ML algorithms were employed to predict the demand for car brands. According to the obtained results, it can be concluded that the SVM had the highest accuracy score compared to other used ML algorithms. The structure of this paper is outlined as follows: Section 2 presents a literature review. In Section 3, we describe the analysis of our dataset, the machine learning techniques used in our study, and the performance metrics employed. Section 4 presents the findings of our study. Finally, Section 5 offers conclusions and recommendations for further research.

2. Literature Review

Predicting future events based on past data is known as forecasting. Demand forecasting is a method that is used to estimate the expected demand for a product or service. As managerial forecasting problems have become more diverse and complex, a variety of forecasting techniques have been developed[4]. There are three main categories of forecasting techniques, namely qualitative, quantitative, and artificial intelligence-based prediction techniques. Quantitative approaches are based on mathematical formulations, while qualitative methods rely more on subjective judgments. Artificial intelligence-based methods can be used to analyze more complex data structures[16]. When data is scarce and the decision-making process is unclear, qualitative methods are usually favored. Surveys, interviews, market analysis, and the Delphi method are the key components of qualitative approaches. Hand observation analysis is considered a quantitative approach, while qualitative techniques depend on statistical techniques such as time series analysis, regression analysis, trend analysis, Holt's and Winter's models, and others. Artificial intelligence (AI) methods utilize sophisticated algorithms that can learn from trial and error. Neural network analysis (NNA), a machine learning technique, is a powerful tool for creating time-series predictions. [17]. Most previous studies have focused on forecasting within the supply chain context, with a heavy reliance on time-series models. These models can detect patterns and trends from historical data and can extrapolate supply chain context into the future, including predicting demand, supply, and prices. Time-series forecasting models can be classified into several categories, such as the Naive approach, moving average methods, exponential smoothing models, ARIMA models, and composite forecasting. Among these, ARIMA is a common statistical technique used to forecast within the supply chain literature.

Wang, Huang, Wang, and Chen [18] proposed a study on predicting inventory demand by combining Taguchi experiments with the ARIMA method and also reviewed the existing literature on ARIMA in the context of supply chain. Gahirwal [19] used HoltWinter and ARIMA techniques to investigate the seasonality and trends in sales data and predict each segment separately. Other statistical methods have also been employed in similar studies. For example, Matsumoto and Ikeda [20] examined the effectiveness of time series models in forecasting the demand for car parts delivered to secondary markets. Additionally, an artificial neural network (ANN) is a popular technique for estimating demand in the supply chain. In recent times, this method has been favored by researchers.

Chawla, Singh, Lamba, Gangwani, and Soni [21] created the ANN model to predict Walmart retail companies' demand in the US. ANN has been used in the supply chain to forecast various types of products. For instance, Ferreira, Martiniano, Ferreira, Ferreira, and Sassi [22] employed it to predict daily product demand, Slimani, Farissi, and Achchab [23] used it to forecast customer product demand in supermarkets, Jebaraj and Iniyan [24]

Doi: https://doi.org/10.54216/FPA.120102
Received: January 09, 2023 Revised: April 08, 2023 Accepted: June 03, 2023

used it for predicting oil demand in India, and so on. They can be found in the literature as hybrid methods that combine various techniques. Jaipuria and Mahapatra [25] combined two methods, Discrete Wavelet Transforms analysis (DWT), and ANN. To evaluate the effectiveness of the suggested approach, they analyzed the ARIMA model. Merkuryeva, Valberga, and Smirnov [26] tested three baseline models for predicting pharmaceutical product demand: the simple moving average, multiple linear regression, and symbolic regression with genetic programming. They found that symbolic regression with genetic programming had the lowest absolute error and mean deviation value and was the most effective method. Fantazzini and Toktamysova [27] developed a multi-dimensional model to forecast the monthly sales of ten different car brands in Germany. They found that Google search data outperformed seasonally adjusted data. Pai and Liu [28] proposed a hybrid model that combines multivariate regression and time series analysis to predict monthly demand for automobile sales using data gathered from Twitter. Their research showed that the hybrid model, which includes tweet sentiment ratings, was more accurate than four-time series models. Table 1 summarizes the differences between conventional and machine learning forecasting methods.

Table 1: Traditional and ML forecasting methods comparison

	ML Forecasting	Traditional Forecasting
Algorithms	Multiple integrated algorithms	A number of single-dimension algorithms
Predictor variables number	Unlimited	Few or single
Source of data	Multiple	mostly historical demand
Requirements of technology	High	Low
Requirements of data	High	Low

3. Methodology

This study aims to respond to the question: Do ML-based forecasting methods in an SC offer more precise predictions of customer demand? Essentially, our goal was to determine whether ML performs better in general. To achieve this goal, we conducted experiments using five different ML algorithms to evaluate their performance in the context of supply chain management. Methods of ML provide features that are ideal for the current forecasting environment. They inherently can process huge datasets and are built to understand patterns from data. The following sections cover in great detail the dataset used, performance indicators, and experimental results.

3.1 Dataset Description

The dataset used is called 'Supply Chain Management for Car' and was collected from Kaggle. It is accessible in a CSV file. Before designing and implementing a model, data understanding and preparation are crucial steps because they provide insight into the data and reveal what corrections or modifications need to be made. After data collection, the dataset was pre-processed to eliminate samples with missing data, remove non-numerical components from non-numerical attributes, convert categorical values into numerical values (if necessary), fix any inconsistencies in the units, and remove attributes that are not relevant to the model's output if necessary. Preliminary data analysis is required to gain a deeper understanding of the data's quality, particularly for descriptive models. To do that, categorical and numerical variables were statistically analyzed, including the identification of key factors that influence the outcomes. This was accomplished by creating a correlation matrix for each attribute to comprehend the relationships between the various components. Some information regarding columns with numeric values was gained as seen in Table 2.

Doi: https://doi.org/10.54216/FPA.120102
Received: January 09, 2023 Revised: April 08, 2023 Accepted: June 03, 2023

Table 2: The dataset columns with numeric values information

	Supplie ID	CarPric	CarModelYe	ProductIE	DateBetweenOrc r	Quantit	Sales	Discour
Max	1000	799454.2	2013	9991.0	345 days	2	999315.6	0.9
Mini	1	500412.4	1953	3.0	-163 days	1	700321.4	0.25
Mean	500.5	649092.1	2000.2	5376.06	90 days	1.5	853098.7	0.57
Std	288.8	85427.2	9.1	3217.9	112 days	0.5	88538.5	0.18
Count	1000	1000	1000	1000	1000	1000	1000	1000

As shown in Table 3, it is evident that the dataset contains more females than males in the gender column. According to Figure 1, females purchased more cars than males. We also wanted to determine which supplier sold the highest quantity of cars. From Figure 2, it is apparent that the supplier named Quartz sold the highest number of cars, with a total of 13 cars sold, followed by Kazu, who sold 12 cars.

Table 3: Number of males and females

Gender	Count
Female	502
Male	498

Customer feedback refers to the information that customers provide regarding their experience with a product or service. The objective is to gain insight into their satisfaction level and help marketing, customer success, and product teams identify opportunities for improvement. Based on our dataset and as shown in Table 4, we observed a higher number of negative (i.e., bad and very bad) feedback than positive (i.e., good and very good) ones.

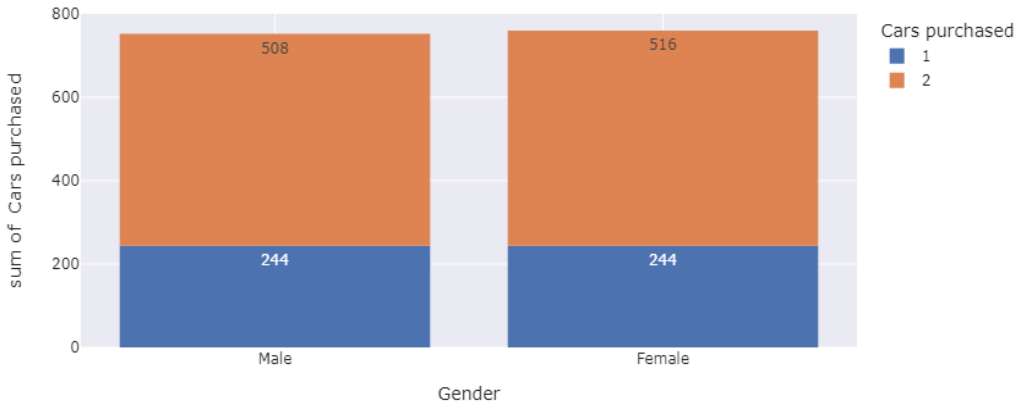

Figure 1: Quantity of cars purchased by female and male

Doi: https://doi.org/10.54216/FPA.120102
Received: January 09, 2023 Revised: April 08, 2023 Accepted: June 03, 2023

The automobile industry places great importance on the value of a strong brand, which can significantly impact consumers' expectations of the price of similar luxury vehicles. Consumers' preferences for different features of mass-market vehicles are influenced by the brand of the vehicle. Premium car brands have a positive impact on all types of cars they produce. Therefore, it is not surprising that automobile manufacturers focus on brand development and positioning as a major marketing priority. Our analysis of the dataset revealed that Chevrolet was the most popular car brand and total revenue of 7,098,858.01 dollars, as illustrated in Figure 3.

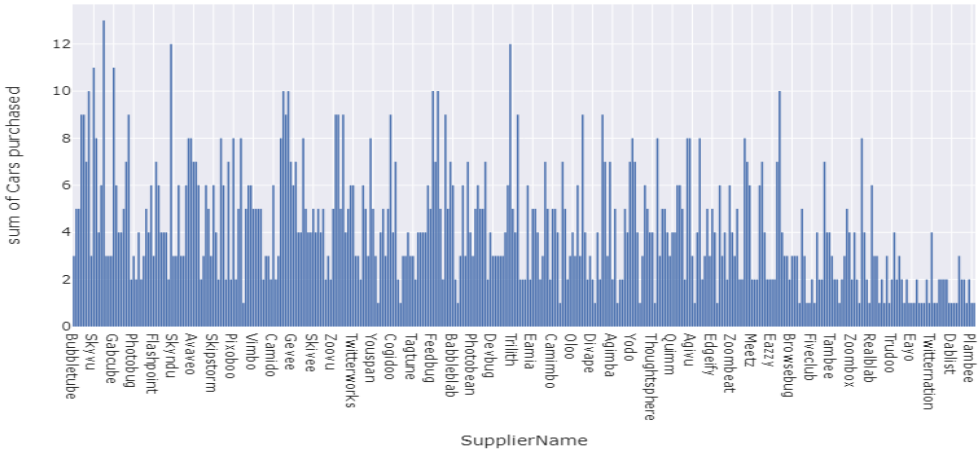

Figure 2: Number of cars sold by each supplier

Figure 4 shows the top twenty suppliers for the Chevrolet car brand in order as per our dataset analysis.

Table 4: Analysis of the dataset customer feedback

Customer feedback	Count
Bad	211
Very Bad	200
Good	186
Very Good	201
Okay	202

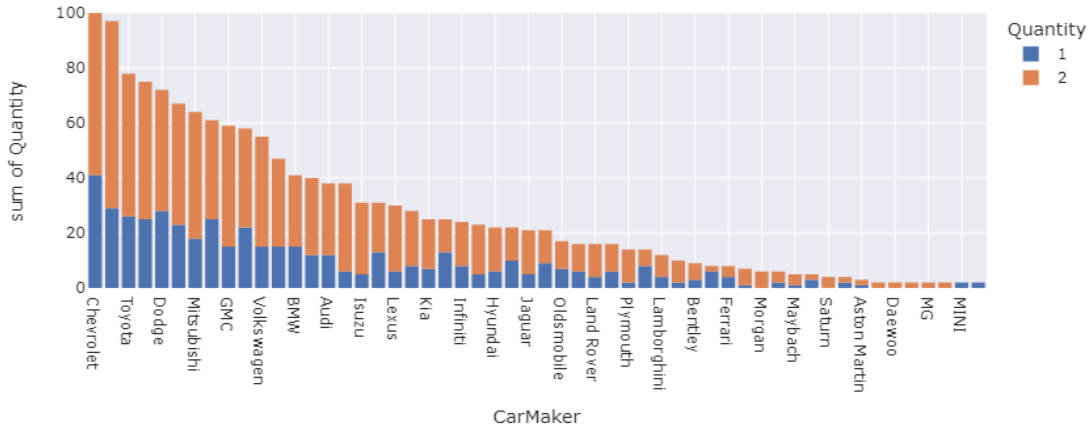

Figure 3: Car brand analysis

Doi: https://doi.org/10.54216/FPA.120102
Received: January 09, 2023 Revised: April 08, 2023 Accepted: June 03, 2023

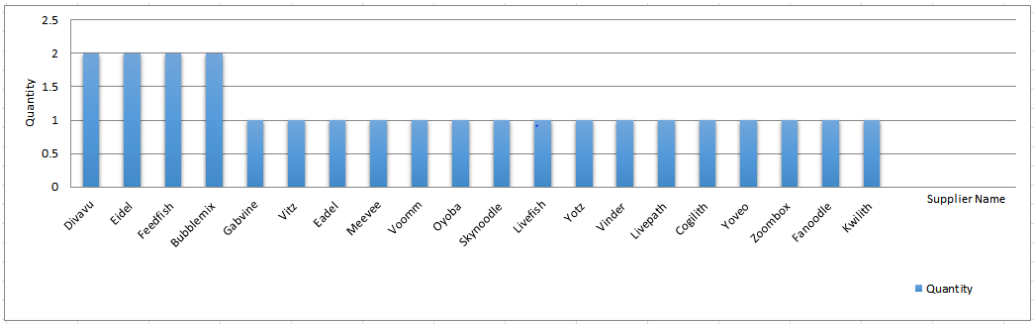

Figure 4: Top twenty suppliers for Chevrolet

3.2 Evaluation Metrics:

To evaluate and appraise the performance of the five proposed ML algorithms, a variety of well-known metrics are produced and defined in terms of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) and defined as:

$$\text{Accuracy} = \frac{TN+TP}{TN+TP+FN+FP} * 100 \quad (1)$$

$$\text{Precision} = \frac{TP}{FP+TP} * 100 \quad (2)$$

$$\text{F1 score} = \frac{2TP}{2TP+FN+FP} * 100 \quad (3)$$

$$\text{Recall} = \frac{TN}{FP+TN} * 100 \quad (4)$$

To examine the correctness and accuracy of the predictive models, we used two additional assessment measures: mean absolute error (MAE) and root mean squared error (RMSE). MAE calculates the average difference between the original and anticipated values, allowing us to evaluate how well the forecasts match the facts. Mathematically, MAE is expressed as:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |yi - y| \quad (5)$$

RMSE, a quadratic scoring rule, calculates the average error magnitude as the average squared difference between the predicted and observed values. Mathematically, RMSE is expressed as:

$$\text{RMSE} = \sqrt{\frac{\sum_{I=1}^{n} (yi - y)^2}{N}} \quad (6)$$

3.3 Used ML Models:

For forecasting demand, a variety of ML techniques are employed. But none of them are equally accurate or precise; for example, one might be inaccurate while another might be more accurate. The ML algorithms used are multiple linear regression(MLR), random forest(RF), k-nearest neighbors(KNN), extreme gradient boosting(XGBoost), and support vector machine(SVM). They have contributed significantly to the evaluation of ML algorithms by offering a much simpler and more comprehensible methodology. Each of these algorithms operates quite differently from the others, with one relying on a specific formula for categorization and prediction and the other creating nodes and trees. The following subsections go into great detail about each applied algorithm.

3.3.1 Multiple Linear Regression(MLR):

MLR is a more complex version of simple regression used to analyze the linear correlation between a dependent variable and multiple independent variables. Once the validity of the dependent variable has been established, independent variables can be identified since they have an impact on the dependent variable. The regression coefficients and constant of each variable are calculated to determine how the independent and dependent variables relate to each other. The MLR model determines the validity of the relationship between the

Doi: https://doi.org/10.54216/FPA.120102
Received: January 09, 2023 Revised: April 08, 2023 Accepted: June 03, 2023

independent and dependent variables using the determination coefficient, R2, which measures how much the dependent variable changes as a result of the independent variables. The R2 value is an indication of the model's goodness of fit[30]. The equation for the regression line takes the following form:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + E \qquad (7)$$

Where:

Y _dependent variable

x_1 , x_2 ,……., x_n _ independent variables

β_0 , β_1 ,……., β_n _ equation parameters

E _ random component (error term of the model)

3.3.2 Support Vector Machine (SVM):

SVM is a popular technique used to improve predictive outcomes. It was introduced by Vapnik as a model for classification and regression tasks and has recently garnered increased interest from data mining, pattern recognition, and machine learning communities due to its ability to generalize well, provide optimal solutions, and possess a strong discriminative capacity. SVM is effective in handling real-world binary classification problems and has outperformed other supervised learning techniques. SVM works by creating decision functions directly from the training data in a higher dimensional space, maximizing the separation (margin) between decision borders. This allows SVM to reduce classification errors in training data and achieve superior generalization capacity compared to other methods, especially when input data is limited. The primary goal of SVM is to partition the training set into different classes using a surface that maximizes the distance between them, thus enhancing the model's generalization capacity[31]. Figure 5 illustrates a scenario where there are many points, and the meeting point is explained.

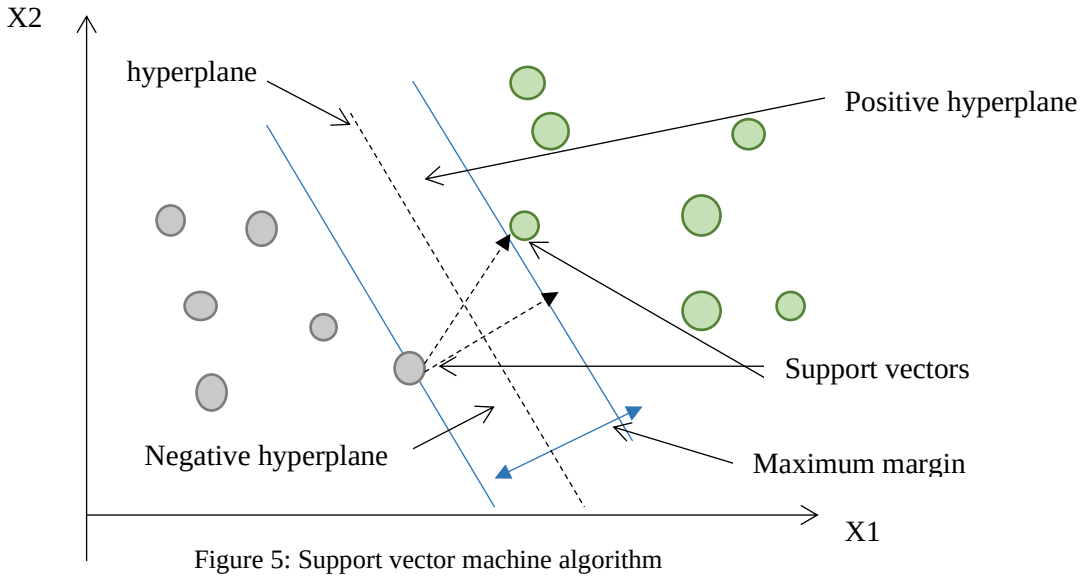

Figure 5: Support vector machine algorithm

Mathematical Constraints On Positive and Negative Data Points:

The positive data points should follow the following restrictions to produce the largest margin between positive and negative data points::

$$W.X_i^+ + b = 1 \qquad (8)$$

The negative data points should also fit:

$$W.X_i^- + b = -1 \qquad (9)$$

Doi: https://doi.org/10.54216/FPA.120102
Received: January 09, 2023 Revised: April 08, 2023 Accepted: June 03, 2023

By employing a clever technique, we can create a uniform equation for both sets of points $Y_i \in \{-1, 1\}$ to indicate the data point's class label X_i :

$$Yi\ (WXi + b) \geq 1 \quad \forall Xi \tag{10}$$

The Hyperplane with Maximum Margin:

$$d_i = \frac{|W\ Xi+b|}{||W||} \tag{11}$$

The quadratic programming problem is given by minimizing the square of the denominator will allow us to maximize this distance:

$$\text{Min } \frac{1}{2} ||W||^2 \quad \text{s.t} \tag{12}$$

$$Y_i\ (WX_i + b) \geq 1 \quad \forall X_i \tag{13}$$

Solution Using The Lagrange Multiplier Method:

The approach of Lagrange multipliers can be used to resolve the quadratic programming problem with the inequality constraints mentioned above. Consequently, the Lagrange function is:

$$LP \equiv \frac{1}{2} ||W||^2 - \sum_{i=1}^{L} \alpha i\ Yi(WXi\ +\ b) + \sum_{i=1}^{L} \alpha i \tag{14}$$

We need to discover the $\boldsymbol{W}$ and $\boldsymbol{b}$ which decreases and the $\boldsymbol{\alpha}$ which increases LP (while keeping $\alpha i \geq 0\ \forall i$). This can be done by differentiating LP with respect to $\boldsymbol{W}$ and $\boldsymbol{b}$ and setting the derivatives to zero:

$$\frac{\partial LP}{\partial w} = 0 \ \rightarrow w = \sum_{i=1}^{L} \alpha i\ YiXi \tag{15}$$

and

$$\frac{\partial LP}{\partial b} = 0 \ \rightarrow 0 = \sum_{i=1}^{L} \alpha i\ Yi \tag{16}$$

The good thing about the previous is that we have an expression for $\boldsymbol{W}$ and $\boldsymbol{b}$ in terms of Lagrange multipliers.

3.3.3 K-nearest neighbors (KNN):

The KNN algorithm is classified as lazy learning, and it selects neighbors from a group of items or objects that share similar traits or characteristics, referred to as the training dataset. The Euclidean distance is used by the algorithm to measure the distance between two points. Prior to using the KNN algorithm, the dataset must be normalized to a standardized scale[32]. The steps for performing the KNN algorithm are depicted in Figure 6. The following formula is used to calculate the Euclidean distance between two points C and D.

$$d(C , D) = \sqrt{(x_2 - x_1^{\ 2}) + (y_2 - y_1)^2} \tag{17}$$

Both classification and regression issues can be solved with KNN, but in the business, it is well recognized for its ability to handle classification ones. The input consists of training samples that are k-closest to the feature space. By examining all the data points, it may identify the group to which a data point belongs[33].

3.3.4 Random Forest (RF)

The RF algorithm is an ensemble method that can be utilized for predictive modeling in both classification and regression problems. The approach involves a random collection of trees that combine to produce the desired outcome. RF follows a collective learning technique where decision trees vote for the most popular class in classification problems or estimate the dependent variable in regression problems. The foundation of this method is decision trees, and a pre-processing step is used to define the nodes of the decision trees that constitute an RF. After creating many trees, the algorithm chooses the best feature from a randomly selected subset of features.

Doi: https://doi.org/10.54216/FPA.120102
Received: January 09, 2023 Revised: April 08, 2023 Accepted: June 03, 2023

Another concept that is created using the decision tree method is generating a decision tree that contributes to the RF. This ensemble of trees is then used to classify new objects based on an input vector[33]. It is known for its ability to make highly accurate predictions and its ability to handle outliers and noise well[34]. It is also capable of calculating error rates more accurately than decision trees. It has been proven that as the number of trees increases, the error rate will always converge[35]. The process of the RF algorithm is shown in Figure 7.

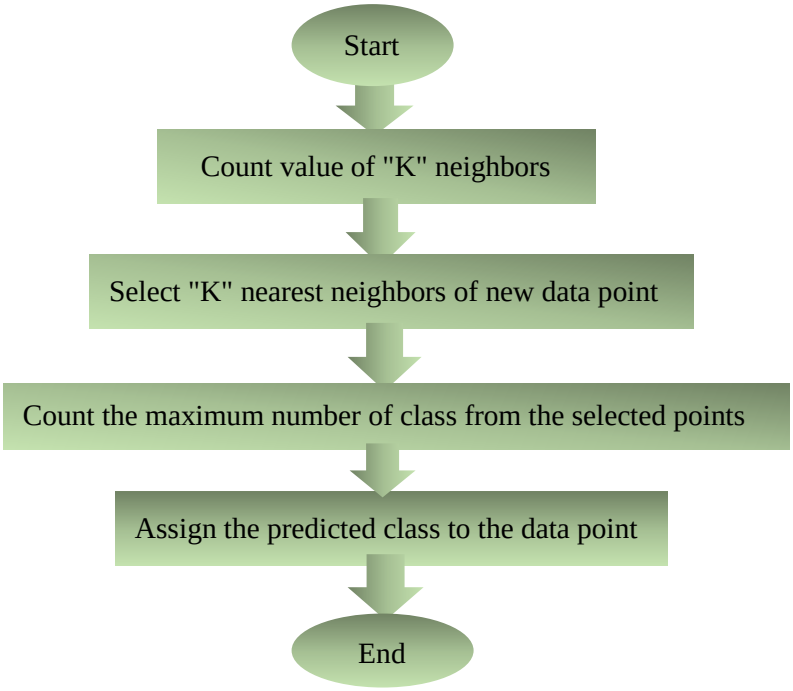

Figure 6: The flowchart of the K-NN algorithm

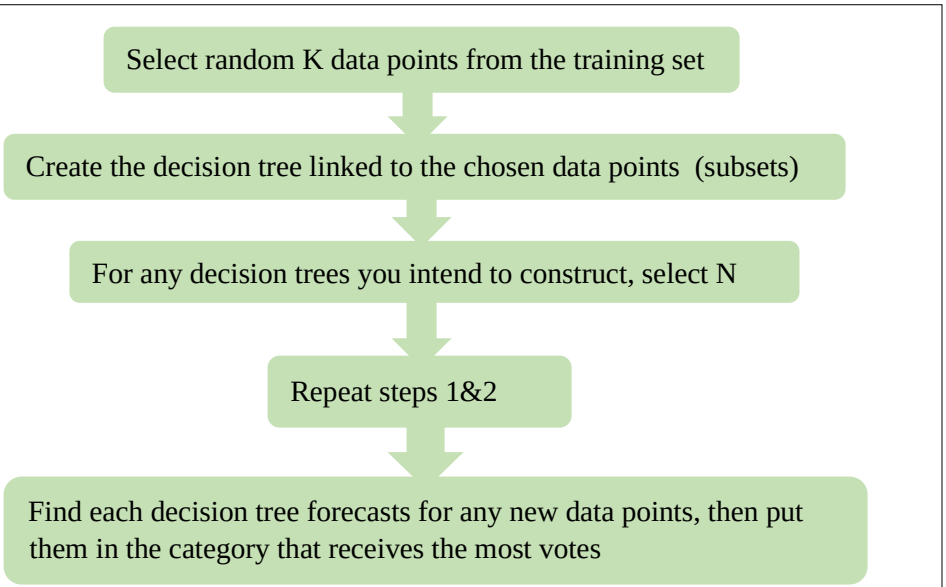

Figure 7: The steps of the Random forest algorithm

Doi: https://doi.org/10.54216/FPA.120102
Received: January 09, 2023 Revised: April 08, 2023 Accepted: June 03, 2023

3.3.5 Extreme Gradient Boosting (XGBoost)

XGBoost is a method that uses trees to classify data, and it has become popular for its high accuracy in recent years [36]. It is a ML algorithm that can handle both classification and regression tasks in a highly scalable and efficient manner [37]. This ensemble method merges weak learners with classification and regression trees (CART) to create a meta-learner that is effective in improving performance [38]. Researchers from various fields have been interested in the XGBoost algorithm due to its remarkable accuracy and speed[39]. It is capable of handling sparse data, and its distributed computing abilities and scalability have produced state-of-the-art results in many benchmark challenges. In different domains, it has been used to address a variety of classification issues. The XGBoost algorithm is combined with the K-means algorithm to forecast short-term load. The primary concept is to calculate the weight of each feature using the XGBoost algorithm to identify features with varying degrees of importance and then specify the weighted distance of the K-means algorithm[40,41]. The algorithm mechanism and its steps are described in detail in Figure 8.

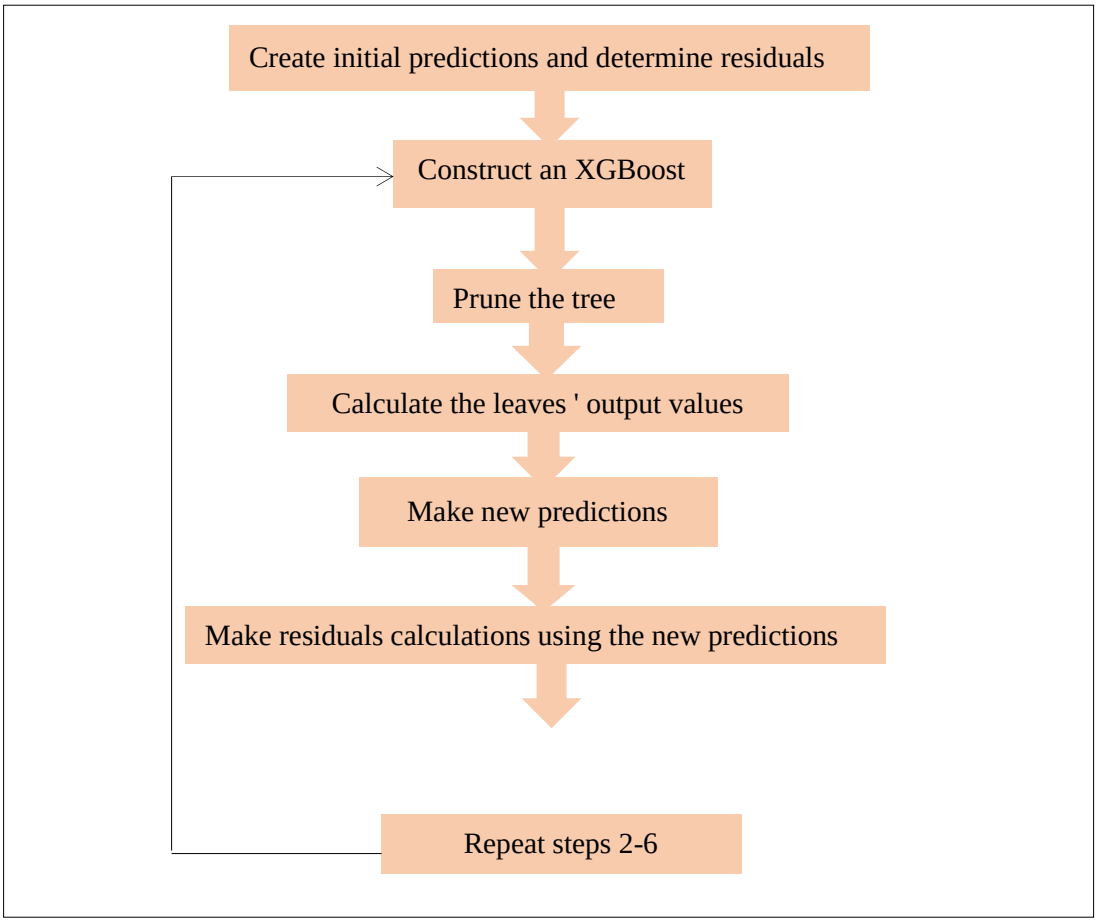

Figure 8: XGBoost algorithm mechanism

4. Experemental Results

ML investigates methods for using computers to learn directly from data and solve problems It is known to have been utilized in forecasting and supplier selection to provide management direction. This is because human judgment errors may occur when making SC judgments, and ML may increase decision-making objectivity. Additionally, SC managers can include ML in their decision-making process. The judgments made by the machine would help in understanding the data from other SC stakeholders, preventing the creation of the bullwhip effect if the entire SC was supported by a machine. These ML applications have demonstrated how

Doi: https://doi.org/10.54216/FPA.120102
Received: January 09, 2023 Revised: April 08, 2023 Accepted: June 03, 2023

urgently forecasting, early warning, and real-time disruption monitoring decision-support systems are required in SCM We compared state-of-the-art ML models on the used dataset to examine their performance. To analyze the performance of models, a train-test split of 75/25 was used in all experiments. Figure 9 displays a graphical illustration of the various performance metrics used by the models. Each experiment was carried out in real-time on the Google Colab platform. Algorithms in Google Colab required less training time than the integrated Jupyter Notebook and Anaconda platform.

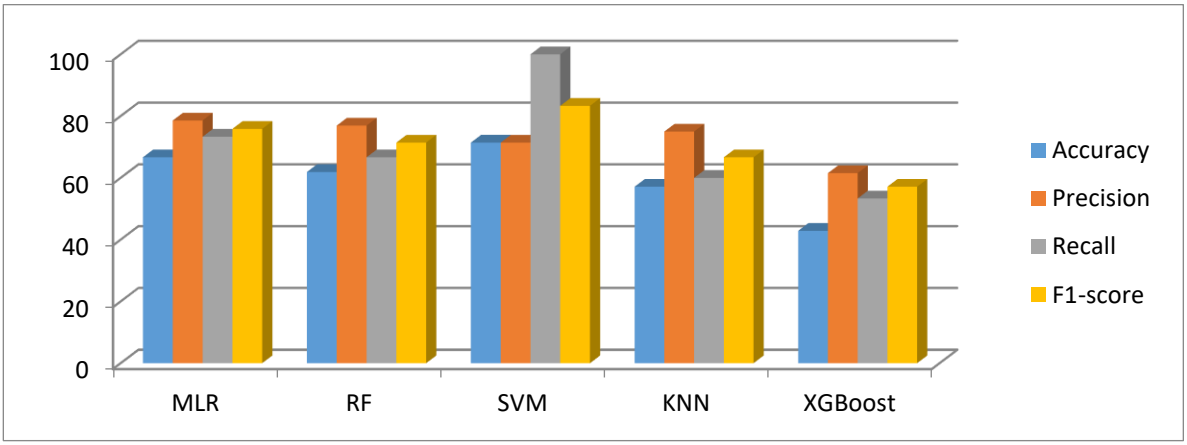

Figure 9: Graphical illustration of the used models' various metrics

Table 5 displays the experimental results for the applied models. As shown in the table, SVM provides the highest accuracy with a value of 71.4285, followed by MLR with an accuracy value of 66.6666. However, the XGBoost algorithm provides the lowest accuracy score of 42.85%. Table 6 lists additional evaluation metrics used in our study.

Table 5: comparison of used ML algorithms' numerical results

No.	Algorithm	Accuracy	Precision	Recall	F1-score
1	MLR	66.6666	0.7857	0.7333	0.7586
2	RF	61.9047	0.7692	0.6666	0.7142
3	SVM	71.4285	0.7142	1.0000	0.8333
4	KNN	57.14	0.75	0.6	0.6666
5	XGBoost	42.8571	0.6153	0.5333	0.5714

Table 6: Various models evaluation metrics for demand forecasting

No.	Algorithm	MAE	RMSE
1	MLR	0.3333	0.5773
2	RF	0.6172	0.3809
3	SVM	0.2857	0.5345
4	KNN	0.4285	0.6546
5	XGBoost	0.5714	0.7559

Doi: https://doi.org/10.54216/FPA.120102
Received: January 09, 2023 Revised: April 08, 2023 Accepted: June 03, 2023

5. Conclusion and Discussion

The main focus of this study is to evaluate and compare the effectiveness of various ML algorithms including KNN, RF, XGBoost, MLR, and SVM, for demand forecasting in supply chain (SC) management. The aim is to determine the strengths and practicality of using ML techniques for this particular task. While experts have access to a variety of forecasting algorithms, there aren't many impartial, trustworthy recommendations regarding which strategy should be employed. We can conclude that ML algorithms could offer demand forecasting with greater accuracy and at a lower cost. The data preprocessing phase of the proposed ML techniques is an important stage that facilitates the generation of inputs to the models. The format of the data in ML projects must be correct to get better results from the applied model. For instance, because the RF approach forbids null values, null values from the initial raw data set must be handled to run the algorithm. Some ML models require more information.. Another consideration is that the data set should be organized so that many ML algorithms can run simultaneously and the best one is selected.

References

[1] Fosso Wamba, Samuel, et al. "Big data analytics in logistics and supply chain management." *The International Journal of Logistics Management* 29.2 (2018): 478-484.

[2] Weng, Tingyu, Wenyang Liu, and Jun Xiao. "Supply chain sales forecasting based on lightGBM and LSTM combination model." *Industrial Management & Data Systems* 120.2 (2020): 265-279.

[3] Petropoulos, Fotios, Rob J. Hyndman, and Christoph Bergmeir. "Exploring the sources of uncertainty: Why does bagging for time series forecasting work?." *European Journal of Operational Research* 268.2 (2018): 545-554.

[4] KAYA, Sema KAYAPINAR, and Özal YILDIRIM. "A prediction model for automobile sales in Turkey using deep neural networks." *Endüstri Mühendisliği* 31.1 (2020): 57-74.

[5] Anggraeni, W., Vinarti, R. A., & Kurniawati, Y. D. (2015). Performance comparisons between Arima and Arimax Method in Moslem Kids Clothes Demand Forecasting: Case Study. Procedia Computer Science, 72, 630–637. Dio: https://doi.org/10.1016/j.procs.2015.12.172.

[6] Seyedan, Mahya, and Fereshteh Mafakheri. "Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities." *Journal of Big Data* 7.1 (2020): 1-22.

[7] Kilimci, Zeynep Hilal, et al. "An improved demand forecasting model using deep learning approach and proposed decision integration strategy for supply chain." *Complexity* 2019 (2019).

[8] Guo ZX, Wong WK, Li M. A multivariate intelligent decision-making model for retail sales forecasting. Decis Support Syst. 2013;55(1):247–55. https://doi.org/10.1016/J.DSS.2013.01.026

[9] Wei J-T, Lee M-C, Chen H-K, Wu H-H. Customer relationship management in the hairdressing industry: an application of data mining techniques. Expert Syst Appl. 2013;40(18):7513–8. https://doi.org/10.1016/J.ESWA.2013.07.053.

[10] Masoumi, S. Maryam, Nima Kazemi, and Salwa Hanim Abdul-Rashid. "Sustainable supply chain management in the automotive industry: A process-oriented review." *Sustainability* 11.14 (2019): 3945

[11] Irina V. Pustokhina,Denis A. Pustokhin. "A Comparative Analysis of Traditional Forecasting Methods and Machine Learning Techniques for Sales Prediction in E-commerce." Journal of American Journal of Business and Operations Research, 10 no. 2 (2023): 39-51.

[12] Pirttilä, Miia, et al. "Working capital management in the Russian automotive industry supply chain." *International Journal of Production Economics* 221 (2020): 107474.

[13] Feizabadi, Javad. "Machine learning demand forecasting and supply chain performance." *International Journal of Logistics Research and Applications* 25.2 (2022): 119-142.

[14] Kourentzes, Nikolaos, Devon Barrow, and Fotios Petropoulos. "Another look at forecast selection and combination: Evidence from forecast pooling." *International Journal of Production Economics* 209 (2019): 226-235.

[15] Huber, Jakob, and Heiner Stuckenschmidt. "Daily retail demand forecasting using machine learning with emphasis on calendric special days." *International Journal of Forecasting* 36.4 (2020): 1420-1438.

[16] Abolghasemi, Mahdi, et al. "Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion." *Computers & Industrial Engineering* 142 (2020): 106380.

Doi: https://doi.org/10.54216/FPA.120102
Received: January 09, 2023 Revised: April 08, 2023 Accepted: June 03, 2023

[17] Makridakis, Spyros, Evangelos Spiliotis, and Vassilios Assimakopoulos. "The M4 Competition: 100,000 time series and 61 forecasting methods." *International Journal of Forecasting* 36.1 (2020): 54-74.

[18] Xiong, Zhiying, et al. "Ciprofloxacin-resistant Salmonella enterica serovar Kentucky ST198 in broiler chicken supply chain and patients, China, 2010–2016." *Microorganisms* 8.1 (2020): 140.

[19] Gahirwal, Manisha. "Inter time series sales forecasting." *arXiv preprint arXiv:1303.0117* (2013).

[20] Matsumoto, M. & Ikeda, A. (2015). Examination of demand forecasting by time series analysis for auto parts remanufacturing. Journal of Remanufacturing, 5(1), 1. Doi: https://doi.org/10.1186/s13243-015-0010-y.

[21] Chawla, A., Singh, A., Lamba, A., Gangwani, N. & Soni, U. (2019). Demand Forecasting Using Artificial Neural Networks - A Case Study of American Retail Corporation. In H. Malik, S. Srivastava, Y. R. Sood, & A. Ahmad (Eds.), Applications of Artificial Intelligence Techniques in Engineering (pp. 79– 89). Singapore: Springer.

[22] Ferreira, R. P., Martiniano, A., Ferreira, A., Ferreira, A. & Sassi, R. J. (2016). Study on daily demand forecasting orders using Artificial Neural Network. IEEE Latin America Transactions, 14(3), 1519–1525. Doi: https://doi.org/10.1109/ TLA.2016.7459644.

[23] Slimani, I., Farissi, I. E. & Achchab, S. (2015). Artificial neural networks for demand forecasting: Application using Moroccan supermarket data. 15th International Conference on Intelligent Systems Design and Applications (ISDA), 266–271. Doi: https://doi.org/10.1109/ISDA.2015.7489236.

[24] Jebaraj, S. & Iniyan, S. (2015). Oil demand forecasting for India using artificial neural network. International Journal of Global Energy Issues, 38(4–6), 322–341. Doi: https://doi.org/10.1504/ IJGEI.2015.070280.

[25] Jaipuria, S. & Mahapatra, S. S. (2014). An improved demand forecasting method to reduce bullwhip effect in supply chains. Expert Systems with Applications, 41(5), 2395–2408. Doi: https://doi.org/10.1016/j.eswa.2013.09.038.

[26] Merkuryeva, Galina, Aija Valberga, and Alexander Smirnov. "Demand forecasting in pharmaceutical supply chains: A case study." *Procedia Computer Science* 149 (2019): 3-10.

[27] Fantazzini, D. & Toktamysova, Z. (2015). Forecasting German car sales using Google data and multivariate models. International Journal of Production Economics, 170, 97–135. Doi: https://doi.org/10.1016/j.ijpe.2015.09.010.

[28] Pai, P. & Liu, C. (2018). Predicting Vehicle sales by sentiment analysis of twitter data and stock market values. IEEE Access, 6, 57655–57662. Doi: https://doi.org/10.1109/ACCESS.2018.2873730.

[29] Hamzah A. Alsayadi, El-Sayed M. El-Kenawy, Abdelhameed Ibrahim,Marwa M. Eid, Abdelaziz A. Abdelhamid. (2022). Blog Feedback Prediction based on Ensemble Machine Learning Regression Model: Towards Data Fusion Analysis. Journal of Fusion: Practice and Applications, 9 (1), 38-46..

[30] Ullah, Ehsan Inam, et al. "Hydrological and ecological impacts of run off river scheme; a case study of Ghazi Barotha hydropower project on Indus River, Pakistan." *Heliyon* (2023): e12659.

[31] Cervantes, Jair, et al. "A comprehensive survey on support vector machine classification: Applications, challenges and trends." *Neurocomputing* 408 (2020): 189-215.

[32] Pandey, Amit, and Achin Jain. "Comparative analysis of KNN algorithm using various normalization techniques." *International Journal of Computer Network and Information Security* 11.11 (2017): 36.

[33] Shah, Kanish, et al. "A comparative analysis of logistic regression, random forest and KNN models for the text classification." *Augmented Human Research* 5 (2020): 1-16.

[34] Ma, Wenyuan, et al. "Identifying forest fire driving factors and related impacts in china using random forest algorithm." *Forests* 11.5 (2020): 507.

[35] Schonlau, Matthias, and Rosie Yuyan Zou. "The random forest algorithm for statistical learning." *The Stata Journal* 20.1 (2020): 3-29.

[36] Parashar, Jyoti, and Munishwar Rai. "Breast cancer images classification by clustering of ROI and mapping of features by CNN with XGBOOST learning." *Materials Today: Proceedings* (2020).

[37] Liew, Xin Yu, Nazia Hameed, and Jeremie Clos. "An investigation of XGBoost-based algorithm for breast cancer classification." *Machine Learning with Applications* 6 (2021): 100154.

Doi: https://doi.org/10.54216/FPA.120102
Received: January 09, 2023 Revised: April 08, 2023 Accepted: June 03, 2023

[38] Alajmi, Mahdi S., and Abdullah M. Almeshal. "Predicting the tool wear of a drilling process using novel machine learning XGBoost-SDA." *Materials* 13.21 (2020): 4952.

[39] Afonso, Ricardo, et al. "Assessment of interventions in fuel management zones using remote sensing." *ISPRS International Journal of Geo-Information* 9.9 (2020): 533.Noh, Jiseong, et al. "Gated recurrent unit with genetic algorithm for product demand forecasting in supply chain management." *Mathematics* 8.4 (2020): 565.

[40] Cherif, Iyad Lahsen, and Abdesselem Kortebi. "On using extreme gradient boosting (XGBoost) machine learning algorithm for home network traffic classification." *2019 Wireless Days (WD)*. IEEE, 2019.

[41] Huiting Zheng, Jiabin Yuan, and Long Chen. Short-term load forecasting using emd-lstm neural networks with a xgboost algorithm for feature importance evaluation. Energies, 10(8):1168, 2017

Doi: https://doi.org/10.54216/FPA.120102
Received: January 09, 2023 Revised: April 08, 2023 Accepted: June 03, 2023