ORIGINAL RESEARCH

A Comparative Study of Demand Forecasting Models for a Multi-Channel Retail Company: A Novel Hybrid Machine Learning Approach

Arnab Mitra[1] · Arnav Jain[1] · Avinash Kishore[1] · Pravin Kumar[1]

Abstract

Demand forecasting has been a major concern of operational strategy to manage the inventory and optimize the customer satisfaction level. The researchers have proposed many conventional and advanced forecasting techniques, but no one leads to complete accuracy. Forecasting is equally important in manufacturing as well as retail companies. In this study, the performances of five regression techniques of machine learning, viz. random forest (RF), extreme gradient boosting (XGBoost), gradient boosting, adaptive boosting (AdaBoost), and artificial neural network (ANN) algorithms, are compared with a proposed hybrid (RF-XGBoost-LR) model for sales forecasting of a retail chain considering the various parameters of forecasting accuracy. The weekly sales data of a US-based retail company is considered in the analysis of the forecasts undertaking the attributes affecting the sale such as the temperature of the region and the size of the store. It is observed that the hybrid RF-XGBoost-LR outperformed the other models measured against various metrics of performance. This study may help the industry decision-maker to understand and improve the methods of forecasting.

✉ Pravin Kumar
pravin.dce@gmail.com

Arnab Mitra
arnabmitra490@gmail.com

Arnav Jain
arnavjain17.4@gmail.com

Avinash Kishore
Avikish7@gmail.com

[1] Department of Mechanical Engineering, Delhi Technological University, Delhi 110042, India

1 Introduction

In the present business world, organizations must improve their services in terms of efficiency, reliability, and availability to survive in the market. Sales forecasting and effective demand planning positively affect the performance of a supply chain [1]. The goal of demand planning is to develop a forecasting model that helps decision-makers in the areas of procurement, production, distribution, and sales. Forecasts serve as the basis for action plans carried out by various organizational units at different planning levels [2]. Managers usually predict sales based on their perceptions, intuitions, and experience. This may vary from person to person and it is very hard to continuously get reliable inputs from qualified and experienced managers. As a result, computer networks can assist in decision-making by estimating future sales. Machine learning (ML) can be utilized to create effective sales forecasting models utilizing the vast amount of data and related information [3].

ML techniques have become prevalent across various disciplines due to their ability to address the problems associated with increasingly large and complex datasets [4]. It involves complex algorithms to reveal meaningful patterns in large-scale and diverse datasets, which would be virtually impossible for a well-trained person [5]. Machine learning focuses on inductive inference inducing general models from specific empirical data [6]. In recent years, advancements in this field have been driven primarily by the creation of new algorithms as well as the ongoing burst of data at reduced computational costs [7].

ML methodsempowered by predictive analysis create enhanced customer engagement andforecast demands with better precision and accuracy in comparison to thetraditional demand forecasting methods [8, 9]. ML techniques can handle complicated correlations between so many causalelements having nonlinear relational demand patterns, thereby boosting retailchain performance [10]. The auto-regressive integrated moving average (ARIMA)and auto-regressive integrated moving average with exogenous variables (ARIMAX)approaches are the most often used predictive models for demand forecasting.Recently developed ML algorithms, such as artificial neural networks (ANN), supportvector machine (SVM), and regression trees, have already outperformed thetraditional methods [11]. The main objectives of the study are as follows:

- To explore the machine learning models used for forecasting.
- To compare the select machine learning models and hybrid models for sales forecasting of a US-based retail company.

In this study, several ML models are compared for retail demand forecasting. These models include random forecast (RF), artificial neural network (ANN), gradient boosting (GB), adaptive boosting (AdaBoost), and extreme gradient boosting (XGBoost), and the performance of these models was compared with the proposed hybrid model of the RF, XGBoost, and linear regression (LR). To make an accurate

comparison of the said models, various performance metrics, namely, mean squared error (MSE), R^2 score, and mean absolute error (MAE), were considered. The advantages and limitations of the employed methodologies as well as future options for performance enhancement are explored. The historical sales data of a leading US-based multinational retail company is considered for the forecasting analysis. The company has a large number of retail stores across the globe, specializing in a long range of products fulfilling the day-to-day demands of consumers. The sales data used for forecasting is related to various stores of the company spread across the USA.

The rest parts of the paper are arranged as follows: Sect. 2 represents the literature review, Sect. 3 discusses the research methodology, Sect. 4 represents the case study and result discussion, and Sect. 5 concludes the study.

2 Literature Review

Many advances and changes have been observed in the global business world during the past couple of decades. Some major factors leading to business-related uncertainties include partner activities, consumer behaviour, rival behaviour, evolving technology, and new product development [12]. Because of these uncertainties, the market is becoming complex and competitive and needs contemporary supply chain management [13]. Precise forecasting is essential for the success of supply chains [14]. On the other hand, various endogenous factors concerning the collection and application of field data can make the forecasting techniques extremely difficult and the external factors can also have a detrimental impact on forecast accuracy [15]. Effective demand forecasting can save more than 7% on the annual operating expenditures of a business [16]. Either qualitative or quantitative techniques can be employed for demand forecasting [17]. Executives' consensus, Delphi technique, historical analogies, and market research are used as qualitative demand forecasting. These techniques heavily rely on domain experts' subjective evaluations and lack decision models, which are data-driven. Quantitative methods, such as regression and time-series analytics, tend to be more systematic and dependable [18]. Regression methods, in particular, are concerned with determining the causal relationships between the independent and the dependent variables [19].

The notion of industry 4.0, ML, and artificial intelligence (AI) as an innovative framework is now being applied to supply chain analytics [20, 21]. Strategic planning, ad-hoc reporting, and end-user computation are common in business intelligence and analytics that aid in robust performance evaluation and management [22, 23]. Descriptive analytics is a technique that deals with happenings in the past, while diagnostic and predictive analytics deals with the happenings to be predicted in the future. For mitigating harmful impacts, the module prescriptive analytics comes in the application [24, 25].

2.1 Machine Learning Methods

ML methods have garnered considerable interest from researchers and practitioners in demand forecasting [26]. But, when it comes to the context of supply chain management, these methods have not been investigated properly. Although algorithms for ML are complex, they offer a variety of distinct and flexible demand forecasting models [27]. The effectiveness of different statistical and ML methods is heavily debated, making it difficult to draw gross generalizations regarding their efficacy [28]. Under varied circumstances, each model class may outperform the others [29]. Some of the most prominent ML methods are discussed in brief in the following paragraphs which have been compared with the proposed hybrid model.

Random Forest Punia et al. [30] introduced a hybrid forecasting method that is based on long short-term memory (LSTM) and random forest (RF). This first model utilizes LSTM to map the temporal characteristics of the data and then random forest is used to model the residuals of the LSTM network. The random forest section of the network is of vital significance as it provides a substantial edge in forecasting accuracy due to its ability to predict sudden changes due to holidays, promotions etc.

XGBoost Kang et al. [31] used the XGBoost hybrid model for tourism and trend prediction. They introduced location attributes and the time-lag effect of network search data to propose the hybrid model. The findings suggest that the spatiotemporal XGBoost composite model outperforms single forecasting approaches. There are several modifiable parameters in the XGBoost algorithm, including general, promotion, and learning objective parameters. They adopted the tree model to concentrate on the nonlinear interactions between the Baidu index and the number of tourists. Using the general parameters, the promotion parameters were changed according to the model chosen.

Gradient Boosting Machine Xenochristou et al. [32] measured the influence of the spatial scale on water demand forecasting. Multiple models were trained on UK daily consumption records for different aggregations of consumptions. Three different levels of spatial aggregation were created using properties' postcodes. A gradient boosting model for training on each of the configurations and prediction for water consumption was made for 1 day in the future. The results implied that the amount of spatial aggregation had a substantial influence on forecasting accuracy and errors can be minimized by utilizing additional explanatory variables.

AdaBoost Walker and Jiang [33] observed that the forecasts using AdaBoost are more accurate and reliable than those derived via a more traditional logistic regression method. They have analysed the importance of each predictor in the AdaBoost model to better understand the relative contribution of each factor to the overall predicted outcome. They observed that AdaBoost models are not easily interpretable as regression model coefficients.

ANN Jahangir et al. [34] used rough artificial neural network (R-ANN) approach to forecast plug-in electric vehicles travel behaviour (PEVs-TB) and PEV load on the distribution network. R-ANNs can increase the accuracy of forecasting findings due to their capacity to analyse phenomena with high uncertainty. Furthermore, two training methods are used in this paper—conventional error back propagation (CEBP) and Levenberg–Marquardt—which are specified using first- and second-order derivatives, respectively. In PEVs-TB, the results demonstrated that the Levenberg–Marquardt approach is more accurate.

It is observed that various authors advocated the varying level of performance of the different models of forecasting which vary on a case-to-case basis. It is very difficult to say which the best performing model is. In this study, RF, ANN, GB, XGBoost, AdaBoost, and hybrid models are tested and compared with each other for the sales forecasting of a US-based retail company.

The hybrid network has been developed by the separate training of the random forest and the XGBoost model on 67% of the data. Then a new dataset was created by generating the prediction from both the models for all the data points. This new dataset was passed as input to train the linear regression model and it predicted the final sales values. Some of the most common machine learning models used in forecasting are summarized in Table 1.

2.2 Measurement of Forecasting Accuracy

Chicco et al. [65] used healthcare information to forecast rates of obesity. R^2 was observed to be more accurate and complete than symmetric mean absolute percentage error (SMAPE). The value of R^2 becomes high if the analysis accurately predicts the majority of ground truth entities for every ground truth category taking into account their dispersion. The accuracy of various machine learning/deep learning models is compared using MSE, MAE, mean absolute percentage error (MAPE), root mean squared error (RMSE), and R^2 metrics [64]. Tsoumakas [3] examined the present state of ML algorithms for forecasting food-purchasing habits. It covers essential design concerns for forecasting food sales, such as the temporal granularity of sales figures, the intake factors to use for predicting sales, and the depiction of the sales evaluation function. It also looks at machine learning algorithms for predicting food sales and important measures like MAE, MSE, RMSE, and MASE, for evaluating forecasting accuracy. Ala'raj et al. [66] used Covid infection data to model and forecast Covid-19 outbreaks. They utilized a modified SEIRD (Susceptible, Exposed, Infectious, Recovered, and Dead) dynamic model and ARIMA model for prediction. The model prediction accuracy was estimated by using 5 metrics: AE, MSE, MLSE (maximum likelihood sequence estimation), normalized MAE, and normalized MSE. Ramos et al. [67] examined the efficiency of phase space models and ARIMA regressors as a tool for predictions of retail sales of five different types of women's footwear: boots, booties, flats, sandals, and shoes. RMSE, MAE, and MAPE were used to evaluate the ARIMA model.

Table 1 Applications of machine learning models in forecasting

Forecasting model	Paper	Description/applications of the forecasting models
Random forest	Islam and Amin [35]	Forecasting utilizing distributed random forest (DRF) and gradient boosting machine (GBM) is examined in this study, which results in extremely simple potential backorder decision scenarios
	Mueller [36]	The predictive capabilities of random forest (RF) and conditional inference forest (CF) are measured. This study explores forecasting using RF which is developed using classification and regression trees (CART) and conditional inference tree (CIT)
	Punia et al. [30]	The research introduced a new forecasting approach that integrates deep learning (DL) with RF and long short-term memory (LSTM) networks. The model has many practical implications for retail managers for a more accurate estimation of complex demand patterns
	Li et al. [37]	The ensemble empirical mode decomposition random forest (EEMDRF) model is introduced for estimating daily electricity usage in conventional organizations. This data is then simulated and predicted via the random forest model
	Rao et al. [38]	A two-stage syncretic cost-sensitive random forest (SCSRF) model is proposed to assess the borrowers' credit risk. It is useful in improving classification and forecasting accuracy
	Ni et al. [39]	RF improved grey ideal value approximation (IGIVA), complementary ensemble empirical mode decomposition (CEEMD), particle swarm optimization algorithm based on dynamic inertia factor (DIFPSO), and backpropagation neural network (BPNN) are combined to develop a hybrid forecasting model. The hybrid model is compared with other models in analysing a PV power plant dataset
XGBoost	Wang et al. [40]	A new load forecasting approach based on SampEn variational mode decomposition (SVMD) and XGBoost is described for industrial customers in this paper. To mitigate detrimental impacts, an adaptive SVMD approach is developed and the XGBoost model is utilized in forecasting industrial customers' short-term electricity demand
	Kang, et al. [31]	A spatiotemporal XGBoost composite model for tourist trend prediction is presented by using geographical features and the time-lag of network search data. The spatiotemporal XGBoost hybrid model outperforms single forecasting approaches, like XGBoost or ARIMA
	Yun et al. [41]	A hybrid genetic algorithm-based XGBoost forecasting methodology is developed. This study empirically demonstrates the significance of feature engineering in forecasting the prices of stocks
	Wang et al. [42]	To overcome the challenges of predicting rental bike demand on an hourly basis, a data mining technique is used. It also investigates a feature filtering strategy to remove non-predictive variables and ranks features based on their prediction performance. Several models are considered in the analysis, namely, XGBoost, LR, and SVM

Table 1 (continued)

Forecasting model	Paper	Description/applications of the forecasting models
	Jabeur et al. [43]	For monthly streamflow forecasting, a hybrid model called Gaussian mixed model (GMM-XGBoost) is developed in which GMM was used to cluster streamflow into multiple groups and XGBoost is used for the prediction
	Osman et al. [44]	Six ML methods have been compared to forecast the gold price; two of them include XGBoost and CatBoost. XGBoost performed the best with the highest R^2 score
	Shi et al. [45]	XGBoost, artificial neural network (ANN), and support vector machine (SVM) are the three ML models tested for forecasting the groundwater levels; the models analysed 11 months of previously observed rainfall, temperature, and evaporation data. The R^2 score of the XGBoost model was the highest
AdaBoost	Walker and Jiang [33]	This study outlines a reproducible adaptive boosting (AdaBoost) model implementation which forecasts the demand-driven acquisition. AdaBoost model's forecasting power is compared to that of a more typical model of logistic regression, and the findings reveal that the AdaBoost model outperforms the others
	Zhou and Lai [46]	Three AdaBoost models are investigated for bankruptcy estimating with missing data by merging them with imputation approaches, and each model is tested on two datasets. For enterprises with or without missing financial data, the AdaBoost M1. Decision Trees (ABM1.DT) algorithm with the KNN approach is reliable in predicting bankruptcy
	Barrow and Crone [47]	This research assesses empirical forecast performance and follows a robust and reliable experimental methodology. The existing AdaBoost.RT and AdaBoost.R2 algorithms can be enhanced by modifying the meta-parameters and the developed AdaBoost.BC
	Wang et al. [48]	To improve generalization ability and reduce ensemble cost, a novel sparse AdaBoost is developed as an amalgamation model. The proposed AdaBoost-ESN (echo state network) technique performs well. AdaBoost-ESN reduces fitting error by 38.25% and forecasting error by 54.96%, respectively
	Sidhu et al. [49]	To anticipate irrigation in rice that is drip-irrigated, ML models such as the linear model, SVM, decision tree, AdaBoost, and RF were utilized. For all of the parameters, AdaBoost offered the best results, with the smallest error value and the biggest R^2 number
	Busari and Lim [50]	AdaBoost-GRU was proposed, wherein the GRU (gated recurrent unit) system was developed, wrapped in sklearn, and then boosted using AdaBoost Regressor. Utilizing daily crude oil prices, the developed framework's forecasting power was compared against AdaBoost-LSTM (long short-term memory). The AdaBoost-GRU ensemble model outperforms AdaBoost-LSTM in terms of delivering the least errors

Table 1 (continued)

Forecasting model	Paper	Description/applications of the forecasting models
	Huang et al. [51]	The proposed method BP AdaBoost (backpropagation AdaBoost) was compared in this research against a single BPNN (backpropagation neural network) model and a simple weighted average BPNN model. The mean value of RMSE was reduced by 8.23%
	Sun et al. [52]	In this study, the ability of the AdaBoost model for bankruptcy forecasting, specifically in the construction industry, has been demonstrated. The predictive powers of AdaBoost and several other algorithms were compared. Based on the experimental results, it was observed that AdaBoost had a more predictive ability as compared to the other models
	Heo and Yang [53]	AdaBoost and LSTM are combined to develop a hybrid ML model. MAPE (mean absolute percentage error) and DS (day-specific) assessment criteria comparison results demonstrate that the suggested approach's out-of-sample forecasting performance is superior to the benchmarks for all the time series
Gradient boosting	Sharma et al. [54]	The block-wise methodology is used to create five separate prediction models: GB, GB-PCA, ANN-CG, and ANN-CG-PCA. By merging the forecasts from the four different models, three hybrid forecasts are developed. The results reveal that the combined model performs better than the separate models, with a MAPE improvement of roughly 15%
	Deng et al. [55]	A hybrid model combining GBDT (gradient boosting decision tree) and DE (differential evolution) is designed to detect insider trading utilizing data from pertinent indexes. The proposed technique is best in terms of recognition accuracy when using a 90-day time window
	Xenochristou et al. [32]	A gradient boosting machine is used to forecast water demand 1 to 7 days in advance. For a reduction in group size from 600 to 5 households, the findings indicate an exponential deterioration in forecasting accuracy from a mean absolute percentage error (MAPE) of 3.2 to 17%
	Yoon [56]	This study examined ML algorithms to project real GDP growth using data from the IMF and the BOJ. The gradient boosting model has a higher forecasting accuracy than the RF model, according to the findings
	Gu et al. [57]	This study introduced a hybrid strategy focused on GBDT, correlation analysis, and EWT for forecasting the London Metal Exchange Nickel settlement price. The developed model was evaluated on three sets of financial data, and the findings demonstrate that when compared to the GBDT, it optimized the TIC output by 16.33%

Table 1 (continued)

Forecasting model	Paper	Description/applications of the forecasting models
	Nie et al. [58]	A new energy consumption prediction model is developed to simulate and predict building electrical energy use. The gradient boosting regression tree approach is used in the proposed model to more precisely estimate energy consumption data. The given prediction model is more accurate and faster to compute
Artificial neural network	Güven and Şimşir [59]	ML techniques have been used for sales forecasting in the retail garment business. ANN and SVM were both trained and evaluated. The ANN model without colour detail outperforms other models by delivering lower RMSE for a larger number of datasets
	Yucesan et al. [60]	Several models such as LR, auto-regressive integrated moving average (ARIMA), ANN, and exponential smoothing were tested while also being evaluated. In addition to this, hybrid models, namely, ARIMA-ANN and ARIMA-LR, were created and utilized in this study. The hybrid ARIMA-ANN achieved a MAPE of 0.49 which was the best out of all the models
	Fanoodi et al. [61]	ANN and ARIMA algorithms are used to predict the demand for blood platelet and reduce the supply uncertainties. It was observed that the ANN and the ARIMA models were far more accurate as compared to the baseline model which was used at Zahedan Blood Transfusion Center to calculate demand volatility
	Jahangir et al. [34]	In this study, an AI-based forward feed and recurrent ANN with an LM training module have been developed to predict the PEV-TB. With the help of the designed model, the net economic loss was reduced by approximately \$16 per year per PEV when compared with conventional methods
	Jebaraj et al. [62]	This study showed that the ANN-based model can be utilized for forecasting coal demand in India as this model gave less MPE and higher R^2 scores for most cases
	Zhao and Yue [63]	The multi-objective linear weighted approach and ANN model are utilized to forecast the security in the supply management of the forest industry
	Loureiro et al. [64]	This study looks into how a DL model can be used in the estimation of sales in the fashion industry. The sales prediction obtained via the DL technique has also been compared with shallow techniques such as decision trees, RF, and SVM. DNN outperformed the other methods in part of the performance metrics considered

3 Methodology

In this study, the performances of XGBoost, RF, ANN, gradient boosting, AdaBoost, and the proposed hybrid framework (RF-XGBoost-LR) are compared using several performance metrics, namely, MAE, MSE, and R^2 score (coefficient of determination). XGBoost, RF, gradient boosting, and AdaBoost are ensemble techniques built on top of decision trees, while ANN is a deep learning technique. The framework of a decision tree (DT) comprises a root node (topmost node), internal nodes, and leaf nodes (end nodes). Simple principles are used in DT algorithms to branch out of the root node, passing through internal nodes and eventually ending in the leaves [68]. In this work, Python 3.7.12 was utilized. For data handling, Pandas version 1.1.5 and Numpy version 1.19.5 were used. For model training, XGBoost version 0.90 and Scikit-learn library version 1.0.2 were used.

3.1 Proposed Framework

In this study, a hybrid model of RF-XGBoost-LR is proposed and its performance is compared with other individual models.

Bagging vs Boosting The primary distinction between the approach of bagging and boosting methods is that the former decreases the variance in prediction by generating the additional data for training from the dataset using combinations with repetitions to produce multi-sets of the original data, while the later adjusts the weight of an observation based on the last classification by iteration. Unlike the bagging method, wherein a uniform selection of each sample is made to build a training dataset, the boosting algorithm's likelihood of choosing a particular sample is unequal. Misclassified or inaccurately calculated samples are more likely to be selected when they carry a higher weight. As a result, each new model may focus on samples that have incorrectly been classified by earlier models [69].

Random Forest RF is an ensemble technique in which the results of many regression trees are combined to generate a single prediction. The primary premise is bagging, in which a sample of training data is selected at random and fitted into a regression tree [70]. This randomly selected sample is termed a bootstrap sample and it is chosen with replacement, meaning any previously chosen data point can be chosen again. A bootstrap sample can be made by choosing N data points randomly from the dataset and then substituting them with the data points present in the dataset. There is a 1/N chance of any data point being chosen.

RF is a combination of decision tree estimators $\{h(X, \Theta k), \; k \; = \; 1, \; 2, ...\}$, in which every decision tree is calculated by utilizing the outputs of a random vector $\{\Theta k\}$, which is independently sampled and evenly distributed among all the decision trees present in the forest.

Once the training is complete, the result of the entire set of decision trees on sample X′ is averaged to generate predictions as shown in Eq. (1).

$$\widehat{f} = \frac{1}{k} \sum_{i=1}^{k} h(X', \Theta k) \tag{1}$$

where $\widehat{f}$ is the final prediction and k is the number of decision trees.

XGBoost An abbreviation for 'extreme gradient boosting' is XGBoost with potential improvements upon gradient boosting. XGBoost enhances the performance and is capable of solving problems of real-world scale while making use of a minimum number of resources [71]. XGBoost is a parallel tree model built upon the gradient boosting model. It utilizes the tree ensemble method, which is made up of a series of CART. Although XGBoost consists of various unique characteristics, second-order Taylor expansion and embedded normalization algorithms appear to be similar to GBDT [72]. XGBoost models have the advantage of scaling effectively for different scenarios while requiring fewer resources than existing prediction models. Within XGBoost, parallel and distributed computation speeds up the model learning and allows for more rapid model exploration.

Hybrid (RF-XGBoost-LR) Model RF makes parallel decision trees which help in reducing the overfitting problem. Accuracy improves as a result of the reduction in variance. In RF, individually separate decision trees are used for each of the multiple copies of original training data. Despite its widespread popularity, random forest suffers from conceptual and practical shortcomings. Random forest adaptive learning is inherently poor in terms of minimizing training error. In particular, each tree is learned autonomously. Complementary information from other trees is not fully realized in this kind of training [73]. It results in a reduction in model performance. XGBoost combines multiple weak learners in a sequential method, which iteratively improves model performance.

XGBoost is a boosting technique. It takes advantage of parallel processing and runs the model on several CPU cores. However, it is affected by the renowned overfitting problem in boosting, which also impacts multiple additive regression trees (MARTs). This challenge arises when there are few trees accessible early in the iteration process; as a result, all of the trees impact significantly the model [74]. Overfitting of training data degrades the model's generalization capabilities, resulting in unreliable performance when applied to novel measurements. High variance and low bias estimators are common manifestations of overfitting. The extra complexity may, and frequently does, aid the model's performance on a set of training data, but it inhibits future point prediction [75]. Overfitting gives an overly optimistic impression of prediction results in new data drawn from the underlying population [76].

Hybrid models are combinations of two or more single models of machine learning or soft computing to achieve higher flexibility along with higher capability in contrast to a single model. One of the two entities, prediction, and optimization of

the prediction are often present in a hybrid model for higher accuracy. Mainly, there are two key reasons to develop a hybrid model:

(i)   To eliminate the risk of an unfortunate prediction of a single forecast in some specific conditions.
(ii) To improve upon the performance of the independent models.

A hybrid model is designed to reap the advantages and overcome the shortcomings of the individual models involved [77]. In this research, a hybrid ML model has been proposed within which random forest regressor which is a bagging technique and XGBoost regressor which is a boosting technique have been combined.

A hybrid model has been developed to overcome the shortcomings of both the models, i.e. RF and XGBoost, as shown in Fig. 1. The random forest model addresses the overfitting problem inherent to XGBoost as it can decrease the model variance without increasing the model bias. This implies that the overfitting problem may be observed in the forecast of a single regression tree, but it can be eliminated in the average forecast of multiple regression trees. Random forest model is poor in terms of reducing the training error as multiple regression trees are trained autonomously. XGBoost addresses this shortcoming of random forest by sequentially training decision trees.

In the proposed framework, RF and XGBoost models are trained separately and predictions of both the models are used as input into an LR model. The LR model processes the final output.

The LR equation can be defined by Eq. (2):

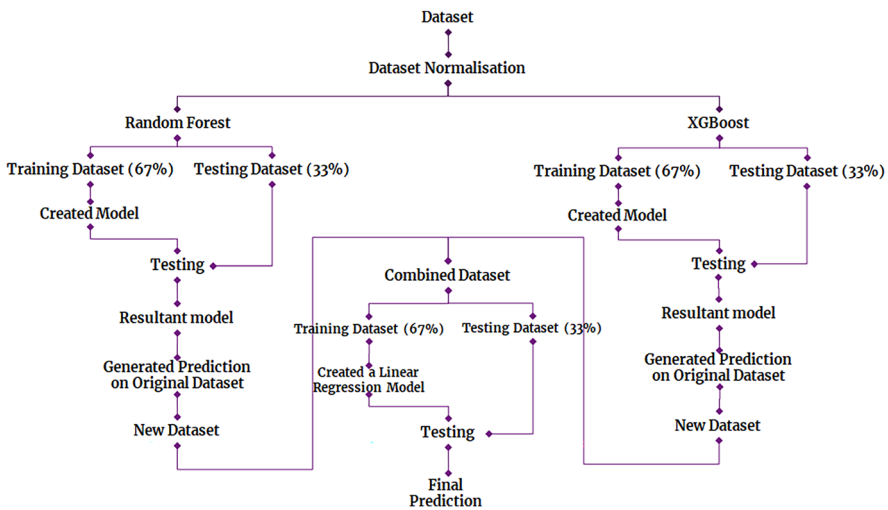$$Y = \beta + \beta_1 \, x_1 + \beta_2 \, x_2 + \varepsilon \tag{2}$$

Fig. 1 Hybrid model flowchart

where the final prediction is represented by *Y*, and the predictions from RF and XGBoost are represented by x_1 and x_2. The *y*-intercept is represented using β. The coefficients and error terms are represented by β_1, β_2, and ε respectively. The value of β_1 is -0.05671126, β_2 is 1.05822592, and β is $-4.4399934967759985e-05$.

The Python source codes for all the four ML forecasting models and hybrid models are shown in the Appendix.

4 Case Study and Result Discussions

The case company operates as a merchandiser of consumer products. The international segment manages supercentres, supermarkets, hypermarkets, warehouse clubs, and cash and carries outside of the USA. The company was founded in 1945 and is headquartered in Bentonville, Arkansas. Among the largest retailers in the world, based in the USA, the company experiences revenue gain year over year. It operates grocery stores, supermarkets, hypermarkets, department stores, and discount stores offering commodities at the lowest prices, the strategy which defines it, in more than 25 countries across the globe. Fuel, gift cards, banking services, and other associated products such as money orders, prepaid cards, and wire transfers are all available through the company.

According to statistics, grocery prices were reduced by an average of 10–15% in markets where the company entered. It has a wide product range, which makes it a tough competitor among other companies in the same segment. Products offered range from electronics and offices, movies, music, and books to jewellery, baby products, and furniture for pharmacies. It is capable of lowering grocery prices by another significant margin during promotional periods. The strong market power over the supplier and competitors allows them to sell the products at the lowest prices and helps them compete in the market.

In this research, the data of a retail company known to keep up with the demands of customers by offering a wide range of products at one stop has been used. The sales data of the company spans different regions in the USA. The data consists of weekly sales for all the 45 stores and 99 departments over 3 years. The data consists of different attributes of the store and geographic-specific information, namely store number, size of the store, department of the store, date mentioning the week, region's average temperature, fuel price in the region, CPI (consumer price index), unemployment rate, and holiday week.

Normalization is a pre-processing step that plays a crucial role in machine learning. Normalizing aids in decreasing the learning time when the datasets are too large. Min–Max normalization transforms the original dataset into the desired interval using a linear transformation. This technique has the advantage of preserving all relations between the data points. Min–Max normalization is given by Eq. (3).

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{3}$$

For proportionate scaling of the data, the Min–Max scale was used at the beginning of the analysis keeping the minimum and maximum values as 0 and 1 respectively.

4.1 Performance Parameters

For the comparison of the forecasting models, mean absolute error (MAE), mean squared error (MSE), and R^2 value are used as discussed in the following subsections.

Mean Absolute Error It is an average measure of errors in a set of predictions. Since it is absolute, it ignores the positivity or negativity of the error and all individual errors are equally weighted. The calculation of MAE is straightforward as shown in Eq. (4). To get the 'total error', the absolute values of the errors are summed up and divided by the total number of observations [78].

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \widehat{y}_i| \tag{4}$$

where γ_i is the true value, $\widehat{\gamma}_i$ is the prediction value, and n is the number of observations.

Mean Squared Error It is also an average measurement of the errors in a set of predictions. The squares of each error are added together and then averaged as shown in Eq. (5). This ensures that all errors are equal in weight and that the direction of the error is irrelevant. Since it is a quadratic function, it will always reach global minima.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 \tag{5}$$

where γ_i is the true value, $\widehat{\gamma}_i$ is the prediction value, and n is the number of observations.

R^2 Score It is also known as the coefficient of determination which expresses the amount of variance in the dependent variable explained by a model [79]. R^2 score is used to evaluate the scattered data about a fitted regression line. Higher R^2 values for similar datasets represent smaller differences between the predicted data and the true data. It measures the relationship between predicted and true data on a scale of 0–1. For example, an R^2 value of 0.8 indicates that the variation of the independent variable explains 80% of the variance of the dependent variable being analysed. It is given by Eq. (6).

$$R^2 = 1 - \frac{SS_{res}}{SS_{total}} = \frac{\sum (y_i - \widehat{y}_i)^2}{\sum (y_i - \mu)} \tag{6}$$

where SS_{res} is the sum of squares of residuals, SS_{total} is the total sum of squares, γ_i is the true value, $\widehat{\gamma}_i$ is the prediction value, and μ is the mean.

R^2 mainly shows whether the said model provides the goodness of fit for the observed values. It was also necessary to understand errors for which metrics namely MSE and MAE were utilized. Mean squared error (MSE)'s purpose was to put more effort into outliers. Due to its square, it weighs large errors more heavily than small ones. Mean absolute error (MAE) is used when measuring the prediction in the same unit as the original series. MAE provides information about how much an average error is expected from the forecast on average. Using the above performance parameters (MSE, MAE, R^2), all the different models incorporated in this study are compared as shown in Table 2.

In the RF model, the number of estimators was kept at 175 and the maximum depth was kept at 28. In the gradient boosting model, the number of estimators was kept at 125 and the maximum depth was kept at 25. In the AdaBoost model, a decision tree with a depth of 25 as a base estimator was used. In ANN, the model has kept 5-layer deep followed by an output layer. The number of neurons in each layer was 10, 12, 24, 12, and 10. The activation function for each layer was taken as 'relu' and 'adam' optimizer. The batch size was taken as 256 and the model was trained for 500 epochs. In the XGBoost model, the number of estimators was kept at 150 and the maximum depth was given as 25. In the RF-XGBoost-LR model, at first, the RF and the XGBoost models were used and their predictions were used as input to an LR model.

In the hyper-parameter optimization phase of the machine learning model, determining the most optimal configuration parameters of the ML optimization methods is challenging. As a result, using random values within the effective range of relevant ML algorithm parameters may result in enhanced optimization outcomes. The output from RF and XGBoost models is being passed as input to an LR model. The LR model was used to output the final predictions because of its simplicity. If we use a complex model like gradient boosting in our final layer, the hybrid model would be prone to overfitting. The hybrid model leads to overcoming the shortcomings of both the RF and XGBoost models.

Wolpert [80] proposed stacking (also known as stacked generalization), which is an ensemble of well-performing models for their capabilities. Stacking uses a single

Table 2 Performance of the forecasting models

S. no	Model	MSE	MAE	R^2
1	Random forest	6.92e-05	0.0028	0.9351
2	XGBoost	4.82e-05	0.0024	0.9547
3	Gradient boosting	9.42e-05	0.0032	0.9116
4	AdaBoost	7.38e-05	0.0027	0.9308
5	Artificial neural network	6.45e-04	0.0129	0.3958
6	RF-XGBoost-LR (hybrid)	**4.79e-05**	**0.0024**	**0.9551**

The bold values shows the performance of the proposed model

model to combine the different predictions from multiple models. Stacked models provide the best results by using a wide range of algorithms in the first layer of design, as different algorithms identify trends and patterns differently in training data, and merging both models results in a more accurate and reliable output.

Various performance measures were utilized to compare the performance of all the models. Holistically, based on three metrics, the proposed forecasting method is found to outperform the other benchmarking methods with an R^2 score of 0.9551, MAE of 0.0024, and MSE of 4.7932e-05.

Figure 2 shows the comparison of the week-wise sale of all the stores and departments against the sales predicted by the hybrid (RF-XGBoost-LR) model. It is observed from Table 2 that the hybrid networks can better forecast the sales as compared to the other models since they can map the trend of the actual sales most accurately.

4.2 Academic Implications

The proposed hybrid model can be utilized to enhance supply chain-related studies and be applied to extend research work on demand forecasting. The robust performance of the proposed framework augments its utility. Retailers, wholesalers, and other industries can use it to their benefit. It is however essential to have adequate domain knowledge for it to be tailored for various applications in different industries. Products in different sectors may have distinct properties that may be retrieved and put into the forecasting framework to fully increase their performance. This makes industry-specific customization to be a potential research topic.

4.3 Managerial Implications

The findings of the study show that the proposed hybrid model improves the forecasting accuracy up to a large extent compared to other individual machine learning
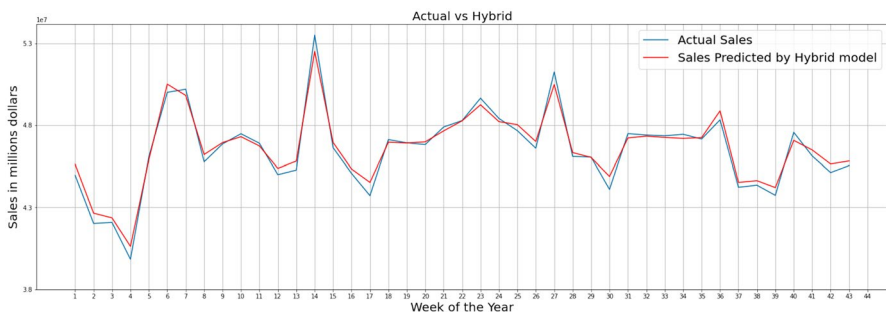

Fig. 2 Comparison between actual sales and forecasting using the hybrid model

models. Both the models random forest and XGBoost jointly overcome the problems of overfitting and training errors in linear regression analysis of the data, and hence, the forecast values are very close to the actual values of sales. Thus, the proposed model helps the industry decision-makers in more accurate forecasting, which leads to the formulation of a better marketing strategy, increasing stock turnover, optimizing capacity building, lowering supply chain costs, and improving customer happiness. An accurate demand forecasting method can improve the supply chain performance by eliminating the bullwhip effect and proper inventory management.

5 Conclusion

In this study, a hybrid model of ML has been proposed combining XGBoost, RF, and LR, for real-time analysis of sales data. Sales data of the retail company with various attributes are trained to introduce a newer more advanced model of veracity. To address the shortcomings of both the RF and XGBoost models, a hybrid model is proposed. At first, the dataset was normalized and then trained and tested separately in RF and XGBoost models. The predictions from these models were assimilated to create a new dataset, which was used as input into the LR model to generate the final predictions.

It is observed that by combining the XGBoost model with the RF model, the dataset improves the accuracy due to reduced variance and enhanced robustness to outliers, which results in improved predictive ability and less vulnerability to overfitting. Three metrics were used in this study, which are MAE, MSE, and R^2 scores. The results suggest that the proposed hybrid model RF-XGBoost-LR (MAE$=0.0024$, MSE$=4.7932$e-05, and R^2 score$=0.9551$) has better performance than the other models, namely RF, ANN, gradient boosting, AdaBoost, and XGBoost. R-squared score infers that the model explains 95.51% of the data and variables incorporated.

A precise demand forecasting in an integrated commercial planning environment can be utilized to optimize capacity building, schedule labour management, inventory, supply chain management etc. In the proposed hybrid model, random forest helps to overcome the overfitting problem of XGBoost, while XGBoost is used to reduce the error by training the decision trees sequentially. Forecasting using the proposed model may improve stock availability and enhance stock allocation.

With all the implications, the proposed hybrid model has some limitations in terms of the requirement of the bid size of the training data, decision integration etc. As the size of training datasets expands, machine learning algorithms become more effective.

Appendix
Python Source Code

```python
#Importing the necessary libraries
from sklearn.ensemble import RandomForestRegressor,GradientBoostingRegressor,AdaBoostRegressor
from sklearn.tree import DecisionTreeRegressor
import sklearn
import xgboost
from sklearn.model_selection import train_test_split,GridSearchCV
from sklearn.metrics import mean_absolute_error,mean_squared_error,r2_score
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import MinMaxScaler
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LinearRegression


#Reading The Dataset
train=pd.read_csv('train.csv')
stores=pd.read_csv('stores.csv')
features=pd.read_csv('features.csv')

#Combining The Data
df = train.merge(features,how="left",on=['Store','Date','IsHoliday']).merge(stores,how='left',on='Store')
df['IsHoliday'] = df['IsHoliday'].astype("int")
df.drop(['MarkDown1','MarkDown2','MarkDown3','MarkDown4','MarkDown5'],axis=1,inplace=True)

#Scaling The Data
X_data,Y_data=np.array(df.drop(['Date','Weekly_Sales'],axis=1).values),np.array(df['Weekly_Sales'].values)
scaler_x = MinMaxScaler()
xscale=scaler_x.fit_transform(X_data)
Y_data=np.reshape(Y_data, (-1,1))
scaler_y = MinMaxScaler()
yscale=scaler_y.fit_transform(Y_data)

#Train Test Split
x_train, x_test, y_train, y_test = train_test_split(xscale,yscale, test_size=0.33, random_state=42)


#Random Forest Model
rf=RandomForestRegressor(n_estimators=175,max_features='auto',max_depth=28)
rf.fit(x_train,y_train)
y_pred_rf=rf.predict(x_test)
print('R2_Score',r2_score(y_test,y_pred_rf))
print('MAE',mean_absolute_error(y_test,y_pred_rf))
print('MSE',mean_squared_error(y_test,y_pred_rf))
y_rf=rf.predict(xscale)
prediction_frame = pd.DataFrame(data=list(y_rf),columns=['RandomForest'])
prediction_frame['Actual']=yscale

#XGBoost Model
xgb=xgboost.XGBRegressor(max_depth=28,n_estimators=150)
xgb.fit(x_train,y_train)
y_pred_xgb=xgb.predict(x_test)
print('R2_Score',r2_score(y_test,y_pred_xgb))
print('MAE',mean_absolute_error(y_test,y_pred_xgb))
print('MSE',mean_squared_error(y_test,y_pred_xgb))
y_xgb=xgb.predict(xscale)
prediction_frame['XGB']=list(y_xgb)


# Creating A DataFrame From All The Predictions
prediction_frame.to_csv('Prediction.csv',index=False)
data=pd.read_csv('Prediction.csv')

# Creating Train And Test Data From The Predictions DataFrame
x_train, x_test, y_train, y_test = train_test_split(data[['RandomForest','XGB']],data['Actual'], test_size=0.33,random_state=42)

#Hybrid Model Using Random Forest and XGBoost Predictions
lr=LinearRegression()
lr.fit(x_train,y_train)
y_pred_lr=lr.predict(x_test)
y_pred_lr=y_pred_lr
print('R2_Score',r2_score(y_test,y_pred_lr))
print('MAE',mean_absolute_error(y_test,y_pred_lr))
print('MSE',mean_squared_error(y_test,y_pred_lr))
```

Data Availability The datasets generated during and/or analysed during the current study are available in the Kaggle.com repository (https://www.kaggle.com/competitions/walmart-recruiting-store-sales-forecasting/data).

Declarations

Competing Interests The authors declare no competing interests.

References

1. Kantasa-Ard A, Nouiri M, Bekrar A, Ait el Cadi A, Sallez Y (2021) Machine learning for demand forecasting in the physical internet: a case study of agricultural products in Thailand. Int J Prod Res 59(24):7491–7515

2. Haberleitner H, Meyr H, Taudes A (2010) Implementation of a demand planning system using advance order information. Int J Prod Econ 128(2):518–526

3. Tsoumakas G (2019) A survey of machine learning techniques for food sales prediction. Artif Intell Rev 52(1):441–447

4. Wilson ZT, Sahinidis NV (2017) The ALAMO approach to machine learning. Comput Chem Eng 106:785–795

5. Goecks J, Jalili V, Heiser LM, Gray JW (2020) How machine learning will transform biomedicine. Cell 181(1):92–101

6. Hüllermeier E (2015) Does machine learning need fuzzy logic? Fuzzy Sets Syst 281:292–299

7. Holzinger A (2016) Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Inf 3(2):119–131

8. Bohanec M, Borštnar MK, Robnik-Šikonja M (2017) Explaining machine learning models in sales predictions. Expert Syst Appl 71:416–428

9. Chase CW Jr (2016) Machine learning is changing demand forecasting. J Bus Forecast 35(4):43

10. Ampazis N (2015) Forecasting demand in supply chain using machine learning algorithms. Int J Artif Life Res (IJALR) 5(1):56–73

11. Smolak K, Kasieczka B, Fialkiewicz W, Rohm W, Siła-Nowicka K, Kopańczyk K (2020) Applying human mobility and water consumption data for short-term water demand forecasting using classical and machine learning models. Urban Water J 17(1):32–42

12. Sillanpää V, Liesiö J (2018) Forecasting replenishment orders in retail: value of modelling low and intermittent consumer demand with distributions. Int J Prod Res 56(12):4168–4185

13. Mohammed A (2020) Towards 'gresilient' supply chain management: a quantitative study. Resour Conserv Recycl 155:104641

14. Oliva R, Watson N (2009) Managing functional biases in organizational forecasts: a case study of consensus forecasting in supply chain planning. Prod Oper Manag 18(2):138–151

15. Van der Laan E, van Dalen J, Rohrmoser M, Simpson R (2016) Demand forecasting and order planning for humanitarian logistics: an empirical assessment. J Oper Manag 45:114–122

16. Van Wassenhove LN, Pedraza Martinez AJ (2012) Using OR to adapt supply chain management best practices to humanitarian logistics. Int Trans Oper Res 19(1–2):307–322

17. Holt CC (2004) Forecasting seasonals and trends by exponentially weighted moving averages. Int J Forecast 20(1):5–10

18. Maia ALS, de Carvalho FDA (2011) Holt's exponential smoothing and neural network models for forecasting interval-valued time series. Int J Forecast 27(3):740–759

19. Wang CH, Chen JY (2019) Demand forecasting and financial estimation considering the interactive dynamics of semiconductor supply-chain companies. Comput Ind Eng 138:106104

20. Jacobs FR, Chase RB, Lummus RR (2014) Operations and supply chain management (pp 533–535). New York, NY: McGraw-Hill/Irwin

21. Stevenson WJ, Hojati M, Cao J (2014) Operations management (p. 182). Chicago-USA: McGraw-Hill Education

22. Lu WM, Wang WK, Lee HL (2013) The relationship between corporate social responsibility and corporate performance: evidence from the US semiconductor industry. Int J Prod Res 51(19):5683–5695

23. Wang CH, Chen YW (2016) Combining balanced scorecard with data envelopment analysis to conduct performance diagnosis for Taiwanese LED manufacturers. Int J Prod Res 54(17):5169–5181

24. Addo-Tenkorang R, Helo PT (2016) Big data applications in operations/supply-chain management: a literature review. Comput Ind Eng 101:528–543

25. Hazen BT, Skipper JB, Ezell JD, Boone CA (2016) Big data and predictive analytics for supply chain sustainability: a theory-driven research agenda. Comput Ind Eng 101:592–598

26. Abolghasemi M, Hyndman RJ, Tarr G, Bergmeir C (2019) Machine learning applications in time series hierarchical forecasting. arXiv preprint arXiv:1912.00370

27. Abolghasemi M, Beh E, Tarr G, Gerlach R (2020) Demand forecasting in supply chain: the impact of demand volatility in the presence of promotion. Comput Ind Eng 142:106380

28. Aye GC, Balcilar M, Gupta R, Majumdar A (2015) Forecasting aggregate retail sales: the case of South Africa. Int J Prod Econ 160:66–79

29. Ahmed NK, Atiya AF, Gayar NE, El-Shishiny H (2010) An empirical comparison of machine learning models for time series forecasting. Economet Rev 29(5–6):594–621

30. Punia S, Nikolopoulos K, Singh SP, Madaan JK, Litsiou K (2020) Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. Int J Prod Res 58(16):4964–4979

31. Kang J, Guo X, Fang L, Wang X, Fan Z (2021) Integration of Internet search data to predict tourism trends using spatial-temporal XGBoost composite model. Int J Geogr Inf Sci 36(2):236–252

32. Xenochristou M, Hutton C, Hofman J, Kapelan Z (2020) Water demand forecasting accuracy and influencing factors at different spatial scales using a gradient boosting machine. Water Resources Res 56(8):e2019WR026304

33. Walker KW, Jiang Z (2019) Application of adaptive boosting (AdaBoost) in demand-driven acquisition (DDA) prediction: a machine-learning approach. J Acad Librariansh 45(3):203–212

34. Jahangir H, Tayarani H, Ahmadian A, Golkar MA, Miret J, Tayarani M, Gao HO (2019) Charging demand of plug-in electric vehicles: forecasting travel behaviour based on a novel rough artificial neural network approach. J Clean Prod 229:1029–1044

35. Islam S, Amin SH (2020) Prediction of probable backorder scenarios in the supply chain using distributed random forest and gradient boosting machine learning techniques. J Big Data 7(1):1–22

36. Mueller SQ (2020) Pre-and within-season attendance forecasting in Major League Baseball: a random forest approach. Appl Econ 52(41):4512–4528

37. Li C, Tao Y, Ao W, Yang S, Bai Y (2018) Improving forecasting accuracy of daily enterprise electricity consumption using a random forest based on ensemble empirical mode decomposition. Energy 165:1220–1227

38. Rao C, Liu M, Goh M, Wen J (2020) 2-stage modified random forest model for credit risk assessment of P2P network lending to "Three Rurals" borrowers. Appl Soft Comput 95:106570

39. Ni L, Wang D, Wu J, Wang Y, Tao Y, Zhang J, Liu J (2020) Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. J Hydrol 586:124901

40. Wang Y, Sun S, Chen X, Zeng X, Kong Y, Chen J, Wang T (2021) Short-term load forecasting of industrial customers based on SVMD and XGBoost. Int J Electr Power Energy Syst 129:106830

41. Yun KK, Yoon SW, Won D (2021) Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process. Expert Syst Appl 186:115716

42. Wang Z, Hong T, Piette MA (2020) Building thermal load prediction through shallow machine learning and deep learning. Appl Energy 263:114683

43. Jabeur SB, Mefteh-Wali S, Viviani JL (2021) Forecasting gold price with the XGBoost algorithm and SHAP interaction values. Ann Operations Res 1–21

44. Osman AIA, Ahmed AN, Chow MF, Huang YF, El-Shafie A (2021) Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. Ain Shams Eng J 12(2):1545–1556

45. Shi R, Xu X, Li J, Li Y (2021) Prediction and analysis of train arrival delay based on XGBoost and Bayesian optimization. Appl Soft Comput 109:107538

46. Zhou L, Lai KK (2017) AdaBoost models for corporate bankruptcy prediction with missing data. Comput Econ 50(1):69–94

47. Barrow DK, Crone SF (2016) A comparison of AdaBoost algorithms for time series forecast combination. Int J Forecast 32(4):1103–1119

48. Wang L, Lv SX, Zeng YR (2018) Effective sparse adaboost method with ESN and FOA for industrial electricity consumption forecasting in China. Energy 155:1013–1031

49. Sidhu RK, Kumar R, Rana PS (2020) Machine learning based crop water demand forecasting using minimum climatological data. Multimed Tools Appl 79(19):13109–13124

50. Busari GA, Lim DH (2021) Crude oil price prediction: a comparison between AdaBoost-LSTM and AdaBoost-GRU for improving forecasting performance. Comput Chem Eng 155:107513

51. Huang H, Zhang Z, Song F (2021) An ensemble-learning-based method for short-term water demand forecasting. Water Resour Manage 35(6):1757–1773

52. Sun S, Wei Y, Wang S (2018) AdaBoost-LSTM ensemble learning for financial time series forecasting. Int Conf Comput Sci (pp 590–597). Springer, Cham

53. Heo J, Yang JY (2014) AdaBoost based bankruptcy forecasting of Korean construction companies. Appl Soft Comput 24:494–499

54. Sharma V, Cali Ü, Sardana B, Kuzlu M, Banga D, Pipattanasomporn M (2021) Data-driven short-term natural gas demand forecasting with machine learning techniques. J Petrol Sci Eng 206:108979

55. Deng S, Wang C, Wang M, Sun Z (2019) A gradient boosting decision tree approach for insider trading identification: an empirical model evaluation of China stock market. Appl Soft Comput 83:105652

56. Yoon J (2021) Forecasting of real GDP growth using machine learning models: gradient boosting and random forest approach. Comput Econ 57(1):247–265

57. Gu Q, Chang Y, Xiong N, Chen L (2021) Forecasting nickel futures price based on the empirical wavelet transform and gradient boosting decision trees. Appl Soft Comput 109:107472

58. Nie P, Roccotelli M, Fanti MP, Ming Z, Li Z (2021) Prediction of home energy consumption based on gradient boosting regression tree. Energy Rep 7:1246–1255

59. Güven İ, Şimşir F (2020) Demand forecasting with color parameter in retail apparel industry using artificial neural networks (ANN) and support vector machines (SVM) methods. Comput Ind Eng 147:106678

60. Yucesan M, Gul M, Celik E (2018) A multi-method patient arrival forecasting outline for hospital emergency departments. Int J Healthcare Manage 13(Sup1):283–295

61. Fanoodi B, Malmir B, Jahantigh FF (2019) Reducing demand uncertainty in the platelet supply chain through artificial neural networks and ARIMA models. Comput Biol Med 113:103415

62. Jebaraj S, Iniyan S, Goic R (2011) Forecasting of coal consumption using an artificial neural network and comparison with various forecasting techniques. Energy Sources Part A Recov Util Environ Effects 33(14):1305–1316

63. Zhao X, Yue S (2021) Analysing and forecasting the security in supply-demand management of Chinese forestry enterprises by linear weighted method and artificial neural network. Enterprise Inf Syst 15(9):1280–1297

64. Loureiro AL, Miguéis VL, da Silva LF (2018) Exploring the use of deep neural networks for sales forecasting in fashion retail. Decis Support Syst 114:81–93

65. Chicco D, Warrens MJ, Jurman G (2021) The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci 7:e623

66. Ala'raj M, Majdalawieh M, Nizamuddin N (2021) Modeling and forecasting of COVID-19 using a hybrid dynamic model based on SEIRD with ARIMA corrections. Infect Dis Model 6:98–111

67. Ramos P, Santos N, Rebelo R (2015) Performance of state space and ARIMA models for consumer retail sales forecasting. Robot Comput Integr Manuf 34:151–163

68. Parsa AB, Movahedi A, Taghipour H, Derrible S, Mohammadian AK (2020) Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. Accid Anal Prev 136:105405

69. Zhang Y, Haghani A (2015) A gradient boosting method to improve travel time prediction. Transport Res Part C Emerg Technol 58:308–324

70. Lahouar A, Slama JBH (2015) Day-ahead load forecast using random forest and expert input selection. Energy Convers Manage 103:1040–1051

71. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system, in proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16). San Francisco, CA, 785–794

72. Kaplan UE, Dagasan Y, Topal E (2021) Mineral grade estimation using gradient boosting regression trees. Int J Min Reclam Environ 35(10):728–742

73. Ren S, Cao X, Wei Y, Sun J (2015) Global refinement of random forest. Proc IEEE Conf Comput Vision Pattern Recogn 723–730

74. Samat A, Li E, Wang W, Liu S, Lin C, Abuduwaili J (2020) Meta-XGBoost for hyperspectral image classification using extended MSER-guided morphological profiles. Remote Sens 12(12):1973

75. Jabbar H, Khan RZ (2015) Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). Computer Sci Commun Instrument Devices 70
76. Steyerberg EW (2019) Overfitting and optimism in prediction models. Clin Predict Models (pp 95–112). Springer, Cham.)
77. Ardabili S, Mosavi A, Várkonyi-Kóczy AR (2019) Advances in machine learning modeling reviewing hybrid and ensemble methods. In International Conference on Global Research and Education (pp 215–227). Springer, Cham
78. Wang Z, Bovik AC (2009) Mean squared error: love it or leave it? A new look at signal fidelity measures. IEEE Signal Process Mag 26(1):98–117
79. Valbuena R, Hernando A, Manzanera JA, Görgens EB, Almeida DR, Silva CA, García-Abril A (2019) Evaluating observed versus predicted forest biomass: R-squared, index of agreement or maximal information coefficient? Eur J Remote Sens 52(1):345–358
80. Wolpert DH (1992) Stacked generalization. Neural Netw 5(2):241–259