



A Novel Approach for Enhanced Feature Selection Over Retail Sales Data Using Ensemble Machine Learning Technique

K. N. Surendra Babu¹ · Mallikarjun M. Kodabagi¹

Received: 27 February 2024 / Accepted: 20 March 2024
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd. 2024

Abstract

The connection between manufacturers and retailers in the retail industry, as well as supply chain management, depend heavily on sales forecasting. The efficiency of traditional methodologies and methods for completing a task has been undermined by the exponential growth of digital data. This study suggests an enhanced feature selection for retail sales using the Citadel POS (Point of Sales) Retail dataset using ensemble machine learning techniques. In order to predict sales data and provide in-depth analysis on retail sales and assessment, a variety of machine learning techniques are used for ensemble sales data. These techniques include diversified regression like Random Forest Regression, Gradient Boosting Regression, Linear Regression, and time series LSTM Model. The information used in this study was given from 2013 to 2019 by Citadel POS, a cloud-based solution that assists retail establishments in managing transactions, inventory, customers, vendors, monitor reports, manage sales, and tender data locally. The proposed method outperformed the regression and time series LSTM models with an MAE of 5.53 and an RMSE of 0.652.

Keywords Retail sales analysis · Ensemble machine learning · Regression models · Time series model · Enhanced feature selection

Introduction

For the IT chain store's inventory management, it is extremely difficult to design an effective sales forecasting model for a variety of reasons, such as under- or over-forecasting, which results in lost sales opportunities and increased operation costs, and over-forecasting, which generates unneeded items [16]. Results from accurate and reliable sales forecasting can improve channel connections, increase customer happiness, and result in significant cost savings. Due to its capacity to capture functional relationships between empirical data, Back Propagation Neural Network (BPN) techniques come in a variety of forms for sales forecasting. However, controlling big parameters and avoiding model over-fitting are two of its drawbacks. The nonlinear

regression estimation problem has been solved using the support vector regression (SVR) algorithm. Because SVR can find a unique solution from the empirical data, its prediction result is superior to that of BPN. However, when numerous possible independent variables are taken into account, SVR is unable to produce accurate findings. The problem can be resolved by using Multivariate Adaptive Regression Splines (MARS). Deo et al. (2017) state that MARS is particularly effective in building models with big datasets, as those for credit scoring, network intrusion detection, and energy price forecasting. Sales forecasting is a crucial tool for businesses to develop business strategies and get a competitive advantage [5]. While there are several time series methodologies that contribute to the subject of sales forecasting, nonlinear data is ignored and only conventional linear data is dealt with, according to Álvarez-Díaz et al. [1]. Many academics are using soft computing approaches, such as fuzzy logic, neural networks, and evolutionary algorithms, to tackle non-linear data issues and provide accurate sales projections in order to circumvent this traditional method.

To handle different problems, a multitude of statistical models and forecasting algorithms have been developed. For instance, the ARIMA model was created to forecast utilising hundreds of previous data points in a couple of seconds [19]. These algorithms, however, cannot handle complicated data patterns that are supplied into a model of forecasting of sales.

This article is part of the topical collection “Advances in Computational Approaches for Image Processing, Wireless Networks, Cloud Applications and Network Security” guest edited by P. Raviraj, Maode Ma and Roopashree H R.

✉ K. N. Surendra Babu
surendrababu.kn@reva.edu.in

Mallikarjun M. Kodabagi
mallikarjun.mk@reva.edu.in

¹ School of Computing and Information Technology, REVA University, Bengaluru, India

In addition to outperforming conventional gradient-based learning algorithms in terms of learning speed, ELM also helps to mitigate a number of drawbacks associated with gradient-based learning techniques. ELM also reduces the learning time of ANNs rapidly [16]. The retailers need to project sales in order to minimise capital costs and make purchasing decisions. It is therefore dependent upon the end users. Therefore, depending on the sort of company, sales forecasting can be done using a statistical model, human planning, or a mix of the two. Through the process of eliminating patterns from past sales data, the proposed approach facilitates the prediction of future sales.

In order to decide which model is most appropriate and successful for the chosen data set, this research will look at the many techniques to sales forecasting used in the financial sector and evaluate the efficacy of the chosen machine learning algorithms. Using Citadel POS data, we have employed time series LSTM model and machine learning based regression models for sales forecasting. The outcomes demonstrated that Xtreme Gradient boosting outperformed other regression strategies as well as time series models.

Preliminaries

We have considered following classifiers in order to carry out our research work.

Naive Bayes (NB)

The NB classifier is employed to determine the probability of an event transpiring in a given scenario. The Bayes probability rule forms its basis. A document is seen by the NB classifier as a bag-of-words, and the likelihood of any given word inside the text is independent of its placement or the presence of other words.

Equation 1 presents the probability of a text document and demonstrates the NB rule. T belongs to class B_k , where $T = \{t_1, t_2, t_3, \dots, t_n\}$ is the collection of feature vectors of the text document, and $B = \{b_1, b_2, \dots, b_k, \dots, b_n\}$ are the output classes for each k items.

$$P(B_k | V) = [P(V | B_k) * P(B_k)] / [P(V)] \quad (1)$$

NB classification produces the posterior probability represented as y in the Eq. 2. The document $t_i \in V$ belonging to class C_k , where $argmax$ denotes the value of the class is mathematically represented by Eq. 2,

$$y = (argmax_y P(C_k) \prod_{i=1}^n P(t_i | C_k)) \quad (2)$$

Decision Tree

The DT predictive model connects an object's experiments to the object's target's result using decision tree learning. The optimal splitting features available at each level are used to construct the tree. In tree structures, class labels are shown as leaves, while feature conjunctions are shown as branches.

Support Vector Machine

Since it usually outperforms the NB, the SVM classifier—also referred to as the non-probabilistic model—is often utilised for text classification. The word vector is generated by converting the classifier review texts. By nonlinearly projecting the training data into higher dimensions, the SVM approach classifies linear data [9]. The hyper planes are used to split the data from the two classes, and a support vector is used to locate the hyper plane. The researchers came to the conclusion that SVM provides a useful technique for classifying documents. Since the SVM is a big margin classifier as opposed to probabilistic classifiers, it outperformed the NB and ME in many cases. Finding the maximal margin hyper plane—which is utilised to divide the document vector inside a class—is the SVM's primary goal.

$$w^{\leftrightarrow} = \alpha_j C_j r^{\leftrightarrow} j \quad (3)$$

In the above Eq. 3, \vec{w} denotes support vector, α_j represents support vectors and C_j is represents the class for \vec{r}_j review vector.

The main contribution of the proposed work is:

The rest of this article is organized as follows. In Sect. "Literature Review" we discuss literature review. Sect. "Proposed Methodology" covers proposed methodology, results and discussion is dealt in Sect. "Results and Discussion" and Sect. "Conclusion" consists of conclusion and future work.

Literature Review

Planning and execution will be the two primary focuses of the supply chain's development.

Forecasting Concept

Future estimations are all that forecasts are. Maybe it's possible to anticipate sunrise and sunset exactly, but that's not how business operates. Because business equations change over time, forecasts could not come to pass. A sales forecast is an estimate of expected demand for the future based on an initial set of environmental conditions, according to Mentzer and Moon [14] (2004). It's important to distinguish between the planning and forecasting procedures. All that planning comprises are the administrative actions required to meet or exceed the sales target. Accurate forecasting aims to precisely anticipate demand. Many

different types of enterprises, governmental organizations, and service sectors have used forecasting. It is often used as a planning input for projects or activities.

[8] et al. enumerates the following features of a sales forecast as below:

- Since forecasts are usually inaccurate, one should constantly anticipate that their errors will be evaluated.
- Generally speaking, short-term projections are more accurate than long-term forecasts. This is due to longer-term projections having a higher standard deviation of inaccuracy compared to the mean.
- Forecasts that are aggregate rather than disaggregated are typically more accurate. Compared to disaggregate forecasts, aggregate forecasts have a lower standard deviation of error.
- The more information is distorted in the supply chain, the more inaccurate the sales prediction is.

Sales Forecasting Need in Planning

Ensuring an adequate supply is the major goal of the manufacturing sector, which aims to meet client demand. In reference to Mentzer and Moon [24] (2004), businesses view sales forecasting as a crucial component of this procedure. To create demand, marketing therefore focuses on end consumers. By employing a number of strategies, including offering services to wholesalers and retailers, the sales department expedites the procedure. Sufficient supply should meet demand. A number of management departments work together to sustain the supply, including buying, production, and logistics.

Forecasting Methods and Techniques

Standardized forecasting methods come in a variety of forms. The forecasting accuracy of each is compared to the degree of quantitative complexity and the logic foundation that forms the basis of the prediction, which informs their differences. Ballon et al. classify such methods into three categories: historical projection, informal, and qualitative.

Using such technology can lead to accuracy for anticipated timeframes. These strategies function best in environments that are stable and have minimal yearly change in the basic demand pattern. [8] According to (Mentzer and Moon, 2004), it is not possible to anticipate all goods using the same time series approach, hence various products require different time series techniques.

Related Work

To precisely predict the actual sales, (Catal et al., 2019) [13] employed a range of machine learning methods. To forecast sales using Walmart's publicly available online sales data, they used a variety of manually constructed time series analysis

techniques and multiple regression algorithms. The R programming language and R packages were utilised in the construction of these techniques.

They used the most precise sales forecast algorithm to develop a web application. Following the trial, the author came to the conclusion that using regression techniques was the best course of action because they performed better than time series analysis processes. MARS (Multivariate Adaptive Regression Splines) is a powerful technique when building models with large datasets, such those for credit scoring, network intrusion detection, and energy price predictions.

(Lu, 2014) concentrated on the previously noted flaws and created a hybrid two-stage model that uses SVR and MARS to accurately forecast sales [12].

(Omar and Liu, 2012) introduced a model based on Back Propagation Neural Network (BPNN) that uses popularity data from magazines discovered via Google Search to improve sales forecasts [17]. As per the author, popular content in magazines has the capability to boost sales. In the proposed methodology, he used popular celebrity names as keywords to get consumers interested in sales forecasting. To determine the popularity of terms, they used a few technologies, such Digg, which allows users to add links in news stories. Using nonlinear historical data, they evaluated our proposed model's forecasting performance. They used information found in Chinese monthly periodicals.

[9] (Holt, 2004) used the Exponentially Weighted Moving Averages (EWMA) technique to calculate the influence on trend of sales. To the seasonal time series sales behavior, they employed two feature cluster-related query techniques. Four models were compared: one that included only the query feature, one that included the seasonal feature and no EWMA model, one that included the seasonal feature and EWMA model, and one that included the seasonal feature and query feature of the recommended model together. The model that was recommended worked the best.

This study provides many feature selection methods and machine learning classifiers to discover the best feature selection of the retail dataset in order to address these issues in the current applications. On the input dataset, pre-processing procedures are first applied. After that, methods like IG and GR are applied to the feature selection procedure. Lastly, we evaluate the suggested method against the dataset using the Linear Regression model, Random Forest Regression, XG Boost, and LSTM Model. When compared to every classifier already in use, the suggested Xtreme Boosting Regression approach performs better.

Proposed Methodology

In order to improve feature selection approaches over retail sales data, an unique strategy is presented in this work. The performance is then assessed using ensemble machine learning

techniques. We are utilizing machine learning techniques to analyze retail sales and the findings of the literature research as input. In this study, we primarily want to assess how well machine learning models such as Xtreme Boosting Regression, Random Forest Regression, and linear regression perform when applied to point-of-sale sales data. The suggested solution's whole technique is depicted in Fig. 1.

Dataset Description

In this paper, early in 2007, we introduced an approach that is being used in a test set of $|SI|=32$ locations for a retail Point of Sale system. Our sales records are all stored in our Citadel Point of Sale system. Using SQL queries, we gathered the data from several tables. The many things for sale are spread across several businesses. There are five stations in every shop. We used just one customer's past information. The client possesses 228 bills. An typical invoice has five items on it. We gathered data between 2013 and 2018, then tested it with 2020 data. Item id, store number, total sales items, and total sales of each item

are all contained in the train data. There are 87,847 rows in the training data set in total.

Data Pre-processing

There are several standardized forecasting techniques available. Their differences lie in how well they foresee compared to the amount of quantitative complexity and the reasoning behind the prognosis. After converting the data into days, weeks, and years, we checked for anomalies and eliminated any missing or null numbers. In order to do testing, we improved the dataset. These techniques might be divided into three categories: qualitative, historical projection, and informal (Ballon, 2004).

a) Augmented Dickey-Fuller Test

An alternative theory proposed by Glynn et al. is that the time series is stationary (2007).

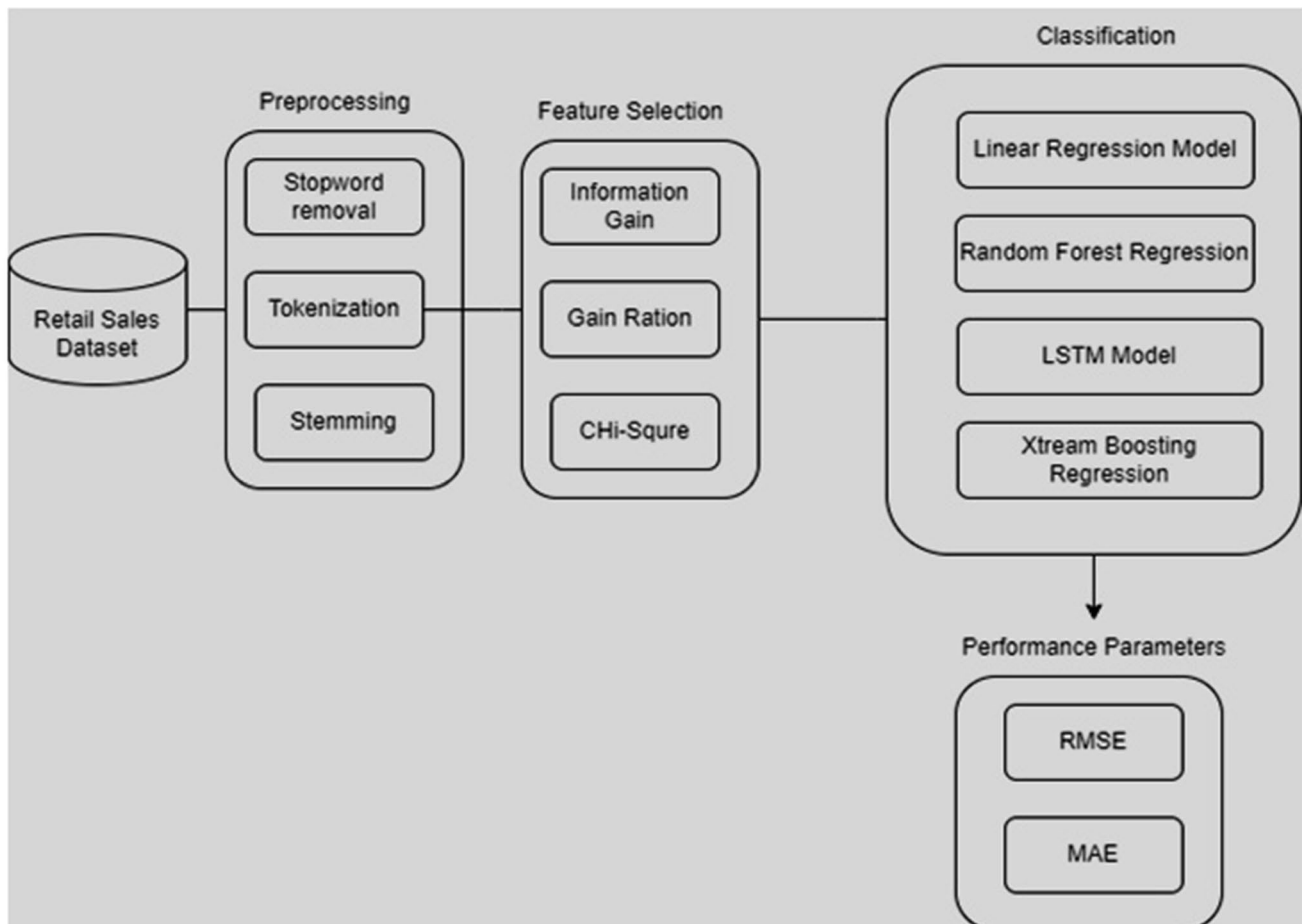
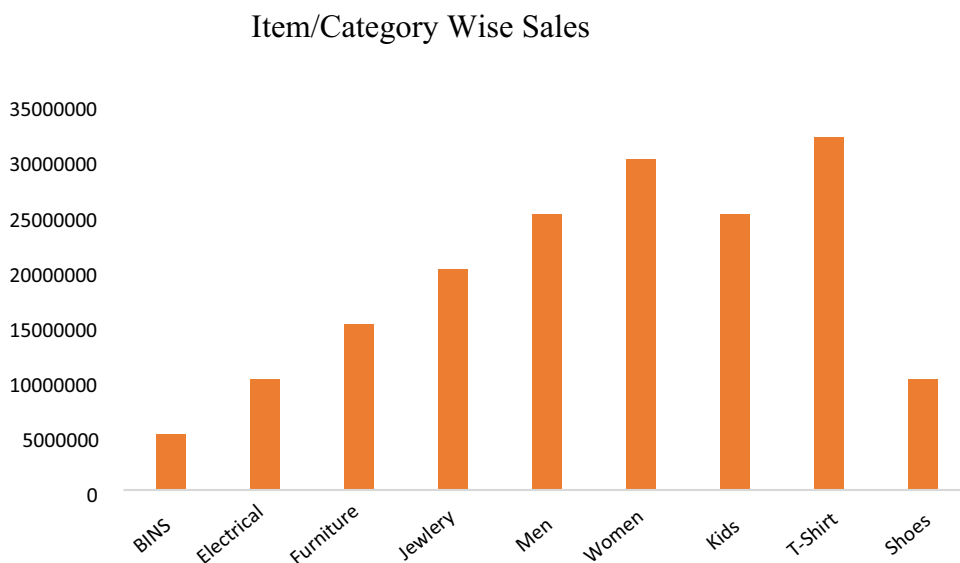


Fig. 1 Proposed System Architecture for Sales Forecasting

Fig. 2 Category/Item Wise Sales Data



Null Hypothesis (H0): If it is not rejected, it indicates that the time series is non-stationary and has a unit root. Its structure varies with time.

Alternate Hypothesis (H1): The time series appears to be stationary since the null hypothesis is rejected, indicating that it lacks a unit root. Its structure is not reliant on time.

The result can be explained as follows:

P-value > 0.05: Since the data has a unit root and is non-stationary, reject the null hypothesis (H0).

P-value <= 0.05: Since there is no unit root in the data and it is stable, reject the null hypothesis (H0).

Feature Selection Techniques

Numerous factors contribute significantly to the effectiveness of machine learning. A crucial component that significantly affects the performance of machine learning models is feature selection. It shortens the training period, decreases overfitting by eliminating redundant data, and raises the model's accuracy. To solve these issues, we employed a variety of strategies, such as the correlation technique. During the feature selection procedure, the feature set exhibiting negative correlations with the target variables was eliminated. In this work, feature selection approaches such as Gain Ratio (GR), Information Gain (IG), and Chi-Square were employed (CHI).

Information Gain(IG)

Entropy is a measure of information that IG uses to activate the qualities required for document review classification [74, 75]. The following defines the estimated amount of data needed to classify the review using training dataset D with m labelled classes:

$$E(D) = \sum_{i=1}^m ((p_i) \log_2(p_i)) \tag{4}$$

In Eq. 4, m denotes the count of classes.

$$D_j.E(D, f_i) = \sum_{j=1}^v (|D_j| \times E(D_j)) \tag{5}$$

In Eq. 5, D_j denotes weightage for j^{th} partition, where $1 \leq j \leq v$ and $E(D)$ representing entropy of the partition.

$$Information-gain(f_i) = E(D) - E(D, f_i) \tag{6}$$

In Eq. 6, information gain denotes information gain on specific feature f_i .

Gain Ratio (GR)

It is working on an iterative process to choose an attribute subspace. [24, 25].

If the split information in the review instances is high, then all of the partitions are homogeneous; if not, the split information is low. The gain ratio for attribute f_i in dataset D is given by Eq. 7.

$$Gain-Ratio = Information - gain(f_i) / Split\ information(f_i) \tag{7}$$

CHI-Squared

In order to determine the link between feature f_i and class C_l , CHI selects the informative features [67]. The ability of feature f_i to categorise the review is exactly proportional to the CHI measure for feature f_i with class C_l [26]. For binary classification with feature f_i , the CHI value is

$$\chi(f_i, C1) = [N(C, f) + N(C, f)] * [N(C, f) + N(C, f)] * [N(C, f) + N(C, f)] \quad (8)$$

In Eq. 8, N denotes the total number of text reviews, N(fi, C1) the number of reviews in class C1 that have feature fi, and N(C1, fi) the number of reviews in class C1 that do not have the feature [27].

Classification

To see our findings, we put the following model into practise and evaluated how well these models performed:

- Gradient Boosting Regression
- LSTM model
- Random Forest Regression
- Linear Regression Model

The MAE is the average of the absolute mistakes, as its name implies (Chai and Draxler, 2014). A lower inaccuracy indicates a higher degree of model correctness.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (9)$$

where yi represents actual values and yi represents the forecasted values.

The square root of the mean square error is known as the root mean square error, or RMSE. A smaller error indicates a higher degree of model accuracy (Chai and Draxler, 2014).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (10)$$

where yi represents actual values and yt represents the forecasted values.

Pseudocode: Algorithm for classification

1. Input: input_dataset
2. Output: Retail Sales data analysis
3. Apply the Pre-processing steps on dataset
 - a. Remove the missing values from the dataset
 - b. Stop word removal
 - c. Tokenization
 - d. Apply Stemming
- # Feature selection techniques
4. IG for each selected features determine entropy using, $E(D) = \sum_{i=1}^m ((p_i) \log_2(p_i))$
5. Determine information gain using entropy, $IG(f_i) = E(D) - E(D, f_i)$
6. Determine gain ratio using split function, Gain-Ratio= Information-gain (fi) / Split information (fi)
7. Based on the informative features calculate chi-Square for the feature fi using, $\chi(f_i, C1)$
- # Classification using proposed ICNN
8. First layer is embedding layer, obtain a word vector The $V \times E$ matrix, is the output of the embedding layer
9. The second layer is a convolutional layer, this layer is the $C_1 \times E$ matrix, stores the local information required to classify the sentiment class in a $V \times E$
10. The third layer is pooling layer, created using max-pooling technique using, $\text{Max}(c) = (\max(c^1), \max(c^2) \dots \max(c^{\text{an}(h+1)}))$
11. Last layer is fully-connected layer, which takes one-dimensional vector as the input, so two-dimensional vector output of pooling layer required to be flattened.
12. Evaluate the proposed Ensemble technique with the performance parameters like classification RMSE, MAE
13. Compare the proposed ICNN with existing Regression Techniques
14. Generate the graphs
15. End

Table 1 Model performance using linear regression

Index	Score
MAE	0.82
RMSE	0.97

Table 2 ARIMA model result

Index	Score
MAE	1.01
RMSE	1.05

Table 3 LSTM model performance results

Index	Score
MAE	0.82
RMSE	0.99

Table 4 Random forest performance results

Index	Score
MAE	0.59
RMSE	0.69

Table 5 Xgboost model performance results

Index	Score
MAE	0.52
RMSE	0.63

Results and Discussion

We must examine both stationary and non-stationary time series in order to determine the outcome. For accurate results, we must thus transform it to a stationary time series.

Citadel POS Dataset

In essence, the Citadel POS is a US-based point of sale system. There are thirty-two sites where various products with varying pricing are for sale. These stores cater to two different kinds of clients: loyalty customers and non-loyalty customers. Regular consumers who shop frequently are considered loyal customers; non-loyalty customers are not regular customers who visit sometimes.

Category/Item Wise Sales Data are displayed in Fig. 2. Stationary data is also verified using the Dickey-Fuller test. We can determine whether the data is stationary by viewing it. Stationary data refers to data whose mean value increases with time. Our data would be stationary if the p value is less than the 5 percent significance level or if the test static value is higher than the crucial value. In this case, the p value is 5.70503.

Predictive Analysis

Linear Regression

In essence, linear regression is a machine learning technique based on supervised learning methodologies. Regression analysis is used to forecast the dependent variable (y) based on the independent value (x).

$$y = m * x + c \quad (11)$$

To begin the prediction work, we first obtained the retail sales information for the years 2013 to 2018. Our model was trained using sales data. We preprocessed the dataset in a few different ways. Following the dataset's training, we verified using a tiny data set that everything was functioning well before moving on to the larger dataset.

Table 1 shows the Mean Absolute Error and Root Mean Squared Error for the validation test derived from Linear Regression. The distance between the data points and the regression line is indicated by the table's standard deviation of the prediction error (RMSE), which is 0.97. The mean absolute error (MAE), on the other hand, is 0.82.

ARIMA Model

The sales are predicted using this methodology. Essentially, it is the time series sales statistical approach. The parameters of the ARIMA model are as follows:

- P: Trend auto regression order. D: Trend difference order.
- Q: Trend moving average order.

Other four differential seasonal factors, such as SARIMA (p, d, q) (P, D, Q) m, can be handled using the SARIMA model and are not included in the ARIMA model.

ARIMA Model Results are displayed in Table 2. RMSE in this table indicates how distant the data points are from the time series problem that is challenging to solve. For instance, forecasting sales to identify trends in data from the stock market. The LSTM model has been used to forecast sales in order to address the sequence issue in the dataset. It is employed to forecast sales using a dataset of retail sales from the past.

Table 3 shows the Root Mean Squared Error and Mean Absolute Error from the validation test's Long Short-Term Memory (LSTM) Regression.

LSTM Model

These kinds of issues have led to numerous sequence prediction problems for a considerable amount of time.

Fig. 3 Linear Regression Sales Forecasting

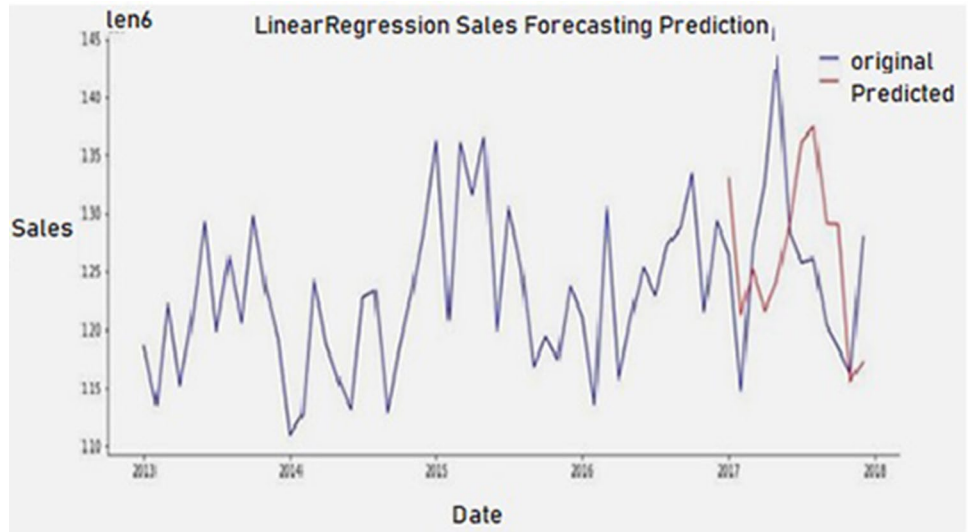


Table 6 Regression model error comparison

Index	RMSE	MAE
Gradient Boosting	0.63	0.52
LSTM	0.99	0.82
ARIMA	1.05	1.01
Linear Regression	0.97	0.82
Random Forest	0.69	0.59

replacement, create many models (decision trees), and then calculate each model's accuracy, and give the model with the highest accuracy more weight.

Table 4 shows the Random Forest Performance Results. The standard deviation of the prediction error (RMSE) in this table, which measures the distance between the regression line and the data points, is 0.69460. The average magnitude of error (MAE), which quantifies the mistake without accounting for the direction differences between the actual and predicted data, is 0.59121.

Random Forest Regression

In order to enhance our findings, we employed the Random Forest regression model. It is employed to increase processing capacity. A decision-tree mechanism is employed in the supervised machine-learning method known as "random forest" to train the model. Using a random training dataset with

Extreme Gradient Boosting Regression

The Xgboost method involves three concepts: boosting, gradient, and extreme. To begin with, boosting is a systematic ensemble strategy that seeks to improve prediction accuracy by turning weak learners—in this case, regression trees

Fig. 4 Model Prediction Comparison

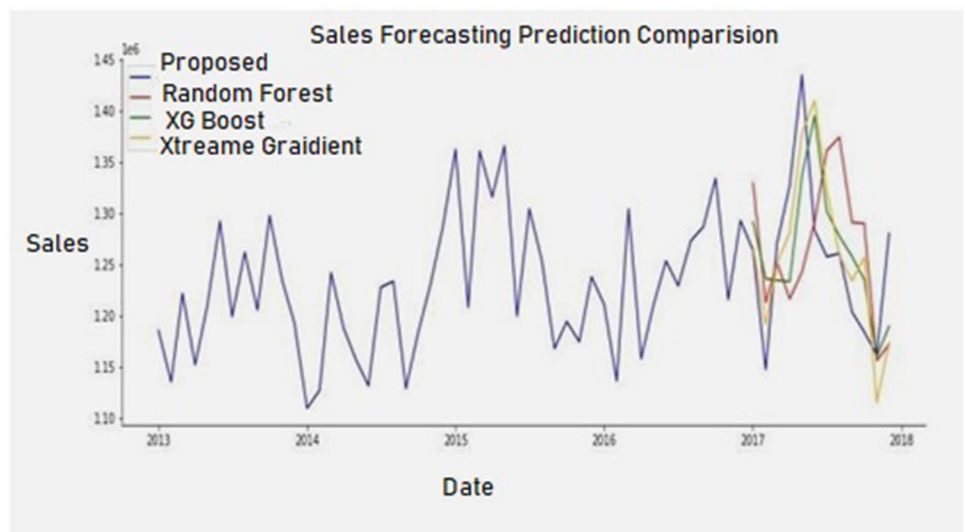
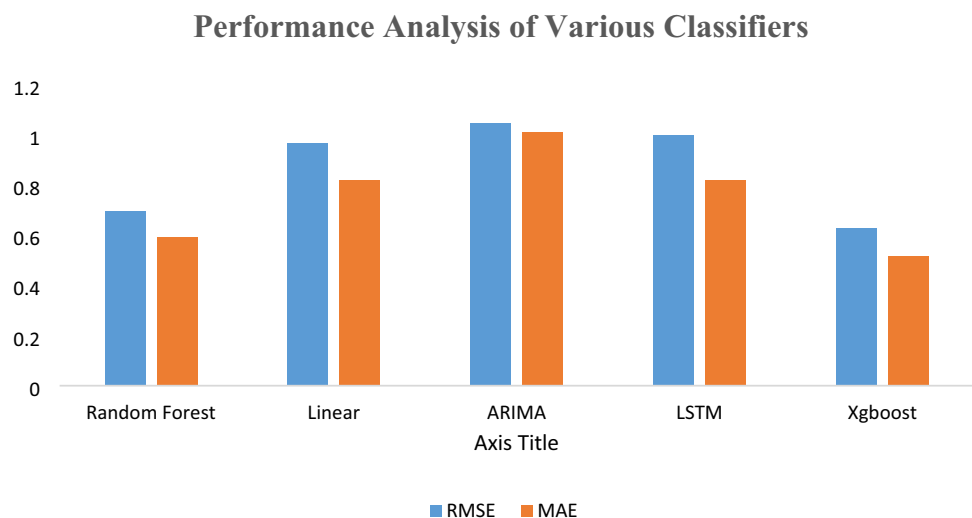


Fig. 5 Comparison Machine Learning Model Error



because the Xgboost model is based on trees; a linear version is also available—into stronger learners. The SMAPE error percentage is 10.14%.

For the validation test, Table 5 shows the Mean Absolute Error and Root Mean Squared Error produced using Gradient Boosting Regression. Without accounting for the directions between the actual and prediction observations, this table displays the MAE, which is 0.52, and the RMSE value is 0.63.

Performance Evaluation and Comparison Results

Several machine learning algorithms were used. We have used various machine learning algorithms to a retail sales dataset. After evaluating each model's performance, we compared them all and came to the conclusion that Xgboost was the most appropriate model for our retail sales dataset.

Figure 3 shows how several model predictions are compared. We used various models on a sales dataset including over 87,746 records. The original value is shown by the blue line, the results of the linear regression are shown by the red line, the results of the Random Forest regression are shown by the green line, and the results of the Xgboost regression are indicated by the orange line. Xgboost is most preferred when compared to all other models.

The performance and outcomes of the models with their default parameter and basic setup are displayed in Table 6. With a greater inaccuracy in both matrices, the ARIMA model performed the poorest.

Figure 4 displays the Comparison Machine Learning Model Error.

Conclusion

According to this research, the most difficult duty for an IT chain store's marketing, customer service, inventory management, and corporate financial planning is sales forecasting. Depending on the type of organization, sales forecasting can be done using statistical models, human planning, or a mix of the two. Accurately creating a sales forecasting model is challenging for several reasons, including over- and under-forecasting. As a result, accurate and successful sales forecasting may lead to a rise in customer satisfaction, improved channel relationships, and significant cost savings. We used time series models like LSTM and ARIMA Model with machine learning regression techniques like Gradient Boosting Regression, Random Forest Model, and Linear Regression Model to estimate sales. We discover that the gradient boosting regression model performs well with the Citadel POS dataset (Fig. 5).

Acknowledgements The authors warmly acknowledged the REVA University, Bengaluru, India for providing the facilities required to carry out the research.

Author Contributions The collaborative efforts and valuable contributions of all authors have been instrumental in advancing the scope and depth of this research endeavor, showcasing their collective dedication and expertise.

Funding No funding received for this research.

Data Availability Upon reasonable request, the corresponding author can provide access to the dataset generated and analyzed in this study.

Declarations

Conflict of Interest No conflict of interest.

References

- Álvarez-díaz, González-gómez, M. & Oterogiráldez, M. S. 2018. Forecasting international tourism demand using a non-linear autoregressive neural network and genetic programming. *Forecasting*, Springer Nature.
- Ballou R. *Business logistics/supply chain management. Planning, organizing and controlling the supply chain*; 2014.
- Catal C, Kaan E, Arslan B, Akbulut A. Benchmarking of regression algorithms and time series analysis techniques for sales forecasting. *Balkan J Electr Comput Eng*. 2019;7:20–6.
- Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE). *Geosci Model Dev Discussions*. 2014;7:1525–34.
- Deo RC, Kisi O, Singh VP. Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. *Atmos Res*. 2017;184:149–75.
- Feng G, Huang G-B, Lin Q, Gay R. Error minimized extreme learning machine with growth of hidden nodes and incremental learning. *IEEE Trans Neural Netw*. 2009;20:1352–7.
- Glynn J, Perera N, Verma R. Unit root tests and structural breaks: a survey with applications, 2007.
- Hofmann E (2013) *Supply Chain Management: Strategy, Planning and Operation*, S. Chopra, P. Meindl. Elsevier Science.
- Holt CC. Forecasting seasonals and trends by exponentially weighted moving averages. *Int J Forecast*. 2004;20:5–10.
- Hussain S, Atallah R, Kamsin A, Hazarika J Classification, clustering and association rule mining in educational datasets using data mining tools: a case study. *Comput Sci On-line Conf*, 2018. Springer, 196211.
- Kaur M, Kang S. Market Basket Analysis: Identify the changing trends of market data using association rule mining. *Proc Comput Sci*. 2016;85:78–85.
- Lu C-J. Sales forecasting of computer products based on variable selection scheme and support vector regression. *Neurocomputing*. 2014;128:491–9.
- Lu C-J, Kao L-J. A clustering-based sales forecasting scheme by using extreme learning machine and ensembling linkage methods with applications to computer server. *Eng Appl Artif Intell*. 2016;55:231–8.
- Mentzer JT, Moon MA (2004) *Sales forecasting management: a demand management approach*, Sage Publications.
- Müller-navarra M, Lessmann S, VOß S Sales forecasting with partial recurrent neural networks: Empirical insights and benchmarking results. In: 2015 48th Hawaii International Conference on System Sciences, 2015. IEEE, 1108–1116.
- Ofoegbu K (2021) A comparison of four machine learning algorithms to predict product sales in a retail store. *Dublin Business School*.
- Omar HA, Liu D-R Enhancing sales forecasting by using neuro networks and the popularity of magazine article titles. 2012 Sixth International Conference on Genetic and Evolutionary Computing, 2012. IEEE, 577–580.
- Pavlyshenko BM. Machine-learning models for sales time series forecasting. *Data*. 2019;4:15.
- Shumway RH, Stoffer DS. *ARIMA models. Time series analysis and its applications*: Springer; 2017.
- Sinaga, K. P. & Yang, M.-S. 2020. Unsupervised K-means clustering algorithm. *IEEE access*, 8, 80716–80727
- Bakhsh M et al. An interpretation of long short-term memory recurrent neural network for approximating roots of polynomials. *IEEE Access* 10 (2022): 28194–28205.
- Tail H, Usman Ashraf M, Alsubhi K, Hani MoaiteqAljahdali The Effect of Fake Reviews on eCommerce During and After Covid-19 Pandemic: SKL-Based Fake Reviews Detection. *IEEE Access* 10 (2022): 2555525564.
- Mumtaz M, Ahmad N, Usman Ashraf M, Alshafut A, Alourani A, Junaid Anjum H Modeling Iteration's Perspectives in Software Engineering. *IEEE Access* 10 (2022): 19333–19347.
- Asif M et al. A novel image encryption technique based on cyclic codes over galois field. *Computational Intelligence and Neuroscience* 2022 (2022).
- Mehak S et al. Automated grading of breast cancer histopathology images using multilayered autoencoder. *CMC-COMPUTERS MATERIALS & CONTINUA* 71.2 (2022): 3407–3423.
- Naqvi MR, Iqbal MW, Ashraf MU, Ahmad S, Soliman AT, Khurram S, Shafiq M, Choi JG. Ontology driven testing strategies for IoT applications. *CMC-Comput Mater Continua*. 2022;70(3):5855–69.
- Tariq S, Ahmad N, Ashraf MU, Alghamdi AM, Alfakeeh AS Measuring the Impact of Scope Changes on Project Plan Using EVM, 8, 2020.
- Asif M, Mairaj S, Saeed Z, Ashraf MU, Jambi K, Zulqarnain RM. A novel image encryption technique based on mobius transformation. *Comput Intell Neurosci*. 2021;17:2021.
- Ashraf MU. A survey on data security in cloud computing using blockchain: challenges, existingstate-of-the-art methods, and future directions. *Lahore Garrison Univ Res J Comput Sci Inform Technol*. 2021;5(3):15–30.
- Ashraf MU, Rehman M, Zahid Q, Naqvi MH, Ilyas I. A survey on emotion detection from text in social media platforms. *Lahore Garrison Univ Res J Comput Sci Inform Technol*. 2021;5(2):48–61.
- Shinan K et al. Machine learning-based botnet detection in software-defined network: a systematic review. *Symmetry* 13.5 (2021): 866.
- Hannan A et al. A decentralized hybrid computing consumer authentication framework for a reliable drone delivery as a service. *Plos one* 16.4 (2021): e0250737.
- Fayyaz S et al. Solution of combined economic emission dispatch problem using improved and chaotic population-based polar bear optimization algorithm. *IEEE Access* 9 (2021): 56152–56167.
- Hirra I, Ahmad M, Hussain A, Ashraf MU, Saeed IA, Qadri SF, Alghamdi AM, Alfakeeh AS. Breast cancer classification from histopathological images using patch-based deep learning modeling. *IEEE Access*. 2021;2(9):24273–87.
- Ashraf MU, Eassa FA, Osterweil LJ, Albeshri AA, Algarni A, Ilyas I. AAP4All: an adaptive auto parallelization of serial code for HPC systems. *Intell Automation Soft Comput*. 2021;30(2):615–39.
- Hafeez T, Umar Saeed SM, Arsalan A, Anwar SM, Ashraf MU, Alsubhi K. EEG in game user analysis: a framework for expertise classification during gameplay. *PLoS ONE*. 2021;16(6): e0246913.
- Siddiqui N, Yousaf F, Murtaza F, Ehatisham-ul-Haq M, Ashraf MU, Alghamdi AM, Alfakeeh AS. A highly nonlinear substitution-box (S-box) design using action of modular group on a projective line over a finite field. *PLoS ONE*. 2020;15(11): e0241890.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.