Journal of Global Research in Electronics and Communication
Volume 1, No. 01, January 2026
Available Online at: www.jgrec.info
RESEARCH PAPER

Predictive Analytics For Sales Using Historical Transaction Data And Seasonal Trends

Mr. Ram Pratap Singh
Department of Computer Science and Engineering
Lakshmi Narain College of Technology,
Bhopal
Email-id - rampratapsps@lnct.ac.in

***Abstract*—The unpredictability of promotional and economic factors makes it imperative to have strong predictive models that would enable sound retail sales forecasting. The accurate sales forecasting in retailing is essential in managing inventory, minimizing operation cost, and maximizing customer satisfaction. To enhance the accuracy of Walmart sales forecasting, the present research utilize a wide range of machine learning (ML) approaches supported by a large amount of data acquired in Kaggle. The methodology includes extensive data pre-processing, such as outlier removal, time-based feature engineering with the assistance of Exploratory Data Analysis (EDA). This is the reason an XGBoost regression model is developed because it has a high possibility to capture nonlinear relationships and can be used with large-scale data. High predictive accuracy is reflected in the results of the experiment with the best model performance as the R 2 of 0.99, Mean Absolute Error (MAE) of 1226.47 and the root mean squared error (RMSE) of 1700.98. A comparative study also proves that XGBoost is better than the classical models such as the Gradient Boosting, Decision Trees and the Random Forests. The results prove the efficiency of the suggested method and indicate its usefulness in practice when applied to retail forecasting.**

Keywords—Sales Forecasting, Predictive Analytics, Retail Sales Data, Seasonal Trends, Demand Forecasting, Machine Learning, Walmart data.

I. INTRODUCTION

In the modern business world, forecasting is one of the main aspects to make decisions and predict the future in organizations. It is not only important to have proper forecasts of demand and sales in terms of efficiency of operation but also competitiveness of operations in unpredictable and volatile markets [1]. Forecasting finds wide usage in the supply chain management, resource allocation, financial planning as well as in marketing strategy [2][3]; therefore, is a subset of a decision-making process by the enterprise level [4][5]. However, its importance, many businesses continue to make subjective choices or qualitative assumptions, which may lead to unproductiveness and opportunity costs. The traditional approaches to forecasting such as time series model and regression model are useful, but deficient in their ability to reflect nonlinearity, seasonality and the multidimensional consumer behavior [6]. These restrictions are the reason why a more advanced data-driven approach is necessary, which could utilize the vast amounts of transactional information at present.

Predictive analytics has emerged as a solution revolution since it uses statistical methodology, machine learning (ML) and artificial intelligence (AI) in the forecasting process [7][8]. Predictive analytics has the potential to assist organizations in identifying the hidden patterns in their past transactions data, seasonal demand patterns, and make more reliable predictions due to historical data and seasonal trends combination [9][10]. This integration enhances the prediction performance and at the same time it makes it easier to optimize the processes, reduce the costs and also improve the satisfaction of the customers [11][12][13]. Recent advancements in ML, in particular, deep learning (DL) images have demonstrated improved results in inferring the complicated temporal and spatial relationships [14]. These models excel in nonlinear association of large data hence, it is superior to the conventional methods in terms of both performance and flexibility [15]. This paper seeks to discuss how predictive analytics can be utilized to predict sales by researching the organization's past history of transactions and seasonal variations in demand. It is intended to demonstrate that a more advanced ML can be used to increase the precision of a forecast, improve the resiliency of the supply chain, and support the planning at the enterprise level. As operational strategies and data-driven insights are interdependent, predictive analytics is a priority to companies in the contemporary dynamic markets that intend to be agile and progressive.

A. Motivation and Contributions of the Study

The motivation behind this research is the fact that the contemporary retail setting requires precise and trustworthy sales forecasts desperately. A proper prediction of the weekly sales is crucial in enhancing inventory control, manpower implementation, and advertisement techniques in big box stores. The Walmart Sales Forecasting Data offers a real-life example that can be used to comprehend how holidays, seasonal trends, and store-level variations can affect the sales trends. Through the creation of a viable forecasting model, companies are able to save their money, prevent cases of stockouts, and make informed decisions that increase the overall productivity of the company. The major findings of this study are the following:

- Robust feature engineering is performed on the Walmart dataset by extracting features to effectively capture temporal sales patterns.
- Trained an XGBoost regressor, a high-performance ensemble method, to accurately predict the continuous target variable, Weekly Sales.
- Assessed model performance using key regression metrics, including R^2, MAE, and RMSE, confirming the framework's reliability and superior predictive accuracy for retail applications.

- The study offers a practical and scalable forecasting framework that supports better retail decision-making during both regular and holiday periods.

The novelty of this work is in its combination of feature engineering based on detailed EDA and a highly optimized XGBoost model to improve the predictions of seasonal sales on the Walmart dataset. The study can use the time-based, lag, and rolling properties in contrast to traditional methods; thus, it takes into account intricate time trends, thereby leading to much better forecasting accuracy. The rationale behind the use of XGBoost can be explained as follows: it can handle large data sets, learn nonlinear associations and use regularization to prevent overfitting. The good experimental outcomes confirm that the suggested approach is effective and applicable in real-world retail forecasting.

B. Organization of the paper

The paper is organised as follows: Section II reviews related work on predictive analytics for sales, Section III describes the Walmart Sales Forecasting dataset, Section IV presents experimental results and evaluation, and Section V concludes the study and makes recommendations for future approaches for retail sales prediction research.

II. Literature Review

This section highlights comparative studies and developments in forecasting models by reviewing relevant publications that have used different ML algorithms for sales prediction.

Rahman Mahin et al. (2025) Numerous ML techniques were used, such as the Elastic Net Regression, ensemble Voting Regressor, Random Forest, KNN, and Linear Regression. RF and KNN perform exceptionally well, however the Voting Regressor outperforms other models because to its strength in various methods. The Voting Regressor has the greatest R^2 of 0.9999 and the lowest RMSE of 1.54. By lowering mistakes and guaranteeing computing efficiency, an ensemble approach increases the accuracy of sales forecasting. The significance of integrating ML into supply chain management is discussed in this study. It demonstrates that the Voting Regressor is the best method for forecasting demand.[16].

Deng et al. (2025) examines the connection between retail market sales trends and commodity attributes by putting forward a strategy for predicting commodity hot-spot categorization based on an ensemble learning model that integrates many ML approaches, such as the GBDT, XGBoost, and LGBM. Additionally, the model classification results are fused using a voting approach. According to the experimental data, the ensemble model outperforms a single model and other methods, with an accuracy of up to 0.91%. This method can help companies find commodities that are in high demand and provide strong support for new commodity designs and adjustments to sales tactics [17].

Balaji and Manikavelan (2025) rates the applicability of some of the ML algorithms including Linear Regression (LR), Import Vector Machine (IVM), and Multi-Scaled Long Short-Term Memory (MLSTM) based on the ability to predict sales of big marts. Though the basic idea of LR was satisfactory, it was deficient in terms of non-linearity and categorical features, which resulted in suboptimal performance on the problem set with a high value of RMSE as 1142.004 and R^2 value as very low at 0.567 on the test data [18].

AbdElminaam et al. (2024) explores the application of Using past data, ML and DL models such as Linear Regression, AdaBoost, Random Forest, XGBoost, SARIMA, ARIMA, and Prophet may forecast future sales. Using two different datasets- Superstore Sales and Walmart Store Sales, the study used to give a complete assessment of the model performance in multiple metrics, R^2 , MAE, RMSE, as well as MAPE. According to results, Linear Regression using AdaBoost works well with the Superstore data with an R^2 score of 77% [19].

Sharif et al. (2024) using Big Mart data and ML models created with PySpark to forecast sales. With an emphasis on XGBoost evaluate and demonstrate the predictive performance of most ML models. Big Mart provides visibility into item type, item weight, and outlet statistics. Consider these attributes to be characteristics of prediction models. PySpark is a powerful distributed computing framework for analysing and training models on massive datasets. Additionally, conduct extensive tests using decision trees (DT), XGBoost, random forests (RF), and linear regressions (LR). Model accuracy improves through training and testing. R-squared and RMSE assess the quality of the model. Metrics assess the model's prediction accuracy and data fit. With an RMSE of 1081 and an R^2 of 0.59, XGBoost demonstrated the best performance [20].

Upadhyay et al. (2023) application of a number of ML methods, including long short-term memory, support vector machines, gradient boosting, random forests, decision trees, linear regression, and convolutional neural networks. A dataset including more than 8,000 sales data entries from various BigMart locations was used to train and assess the models. With an R^2 value of 0.68 and 0.69, the findings demonstrated that the convolutional neural network and long short-term memory models performed more accurately than the other techniques [21].

Although the sales forecasting process has advanced tremendously through the application of ML and DL methods, there are still gaps in the research. Ensemble and hybrid models have been shown to be very accurate in terms of prediction, but most methods continue to have difficulty in non-linear relationships, categorical data and other large-scale heterogenous data sets, which results in poor performance in real life. CNNs and LSTMs are models of DL, which demonstrate better accuracy, but are limited by the size of data and complexity. Moreover, the majority of researches work with the datasets of a single store or region, which restricts the ability to apply the models to wider markets. Table 1 provides a summary of the methodologies, datasets, performance metrics and limitations of the existing research, indicating the necessity of more robust, scalable, and hybrid predictive frameworks with the capacity to handle a range of transactional features and effectively employ ML and DL techniques.

TABLE I. Summary of Recent Machine Learning and Deep Learning Approaches for Sales Forecasting

Reference	Methodology	Dataset	Performance	Limitations & Future Work

Rahman Mahin et al., (2025)	KNN, Random Forest, Voting Regressor (ensemble), Elastic Net Regression, and Linear Regression	Not specified (transaction datasets for sales)	Voting Regressor achieved RMSE = 1.54, R^2 = 0.9999; outperformed other models	Future studies might examine deep learning algorithms to improve prediction powers, use in larger markets, and integration of other aspects
Deng et al., (2025)	Ensemble learning: Gradient Boosting Decision Tree, XGBoost, LightGBM, combined with voting strategy	Retail market sales and commodity features dataset	Accuracy = 0.91; ensemble model outperforms single models	Future work could include application to more diverse commodities and dynamic market conditions
Balaji and Manikavelan, 2025	Linear Regression (LR), Import Vector Machine (IVM), and Multi-Scaled Long Short-Term Memory (MLSTM) models for Big Mart sales prediction	Big Mart Sales Dataset	RMSE = 1142.004; R^2 = 0.567	Linear models underperform on non-linear and categorical features; suggest hybrid deep learning approaches for higher accuracy
AbdElminaam et al., 2024	Comparative analysis of ML and DL models — Linear Regression, AdaBoost, Random Forest, XGBoost, SARIMA, ARIMA, Prophet	Superstore Sales and Walmart Store Sales datasets	Best model (LR + AdaBoost) achieved R^2 = 0.77	Moderate performance; future studies could explore hybrid models integrating time-series and feature-engineered inputs
Sharif et al., 2024	PySpark machine learning models for distributed sales forecasting (XGBoost, Decision Tree, Random Forest, Linear Regression)	Big Mart dataset with item, visibility, type, and outlet attributes	XGBoost: RMSE = 1081; R^2 = 0.59	Limited scalability evaluation; suggests inclusion of deep learning and advanced feature selection methods
Upadhyay et al., 2023	Sales forecasting techniques include SVM, CNN, LSTM, Random Forest, Decision Tree, Gradient Boosting, and Linear Regression	Big Mart Sales dataset (~8000 entries)	CNN: R^2 = 0.68; LSTM: R^2 = 0.69	Performance limited by dataset size; recommends using larger datasets and ensemble deep learning for better prediction accuracy

III. METHODOLOGY

This methodology uses the Walmart sales dataset depicted in Figure 1 to propose a machine learning-based system for sales forecasting. This is followed by data pre-processing, handling missing values, converting dates, removing outliers using Z-scores, and generating time-based and lag features. The refined dataset is then divided into Train data and Test data with an 80:20 ratio. The training data is used to train an XG BOOST model, which is subsequently utilized to predict the test data. The final model's performance is assessed using key evaluation metrics, including R^2, RMSE, MSE, and MAE, providing a reliable solution for accurate retail predictions.

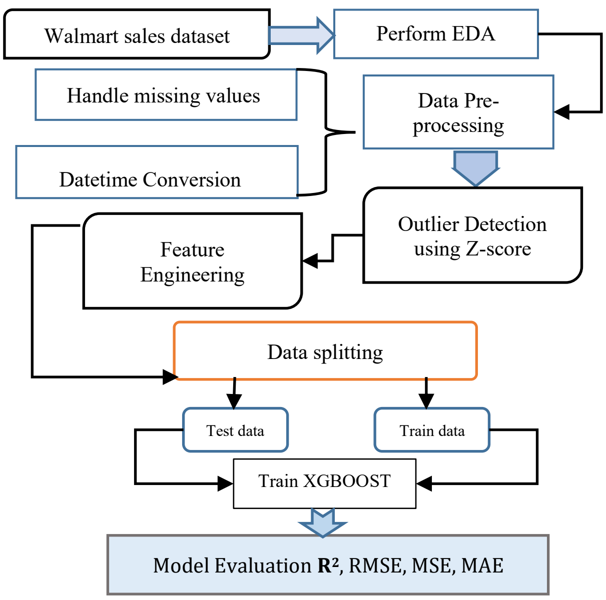

Fig. 1. Proposed flowchart for Walmart Sales data

A. Data Collection

The Walmart Sales Forecasting data would offer the in-depth reference of time-series retail sales prediction. It includes past weekly sales information of 45 Walmart stores in various regions, economic and promotion information. The total number of values in this dataset is 6414 and the number of features is 8, all the features are quantitative.

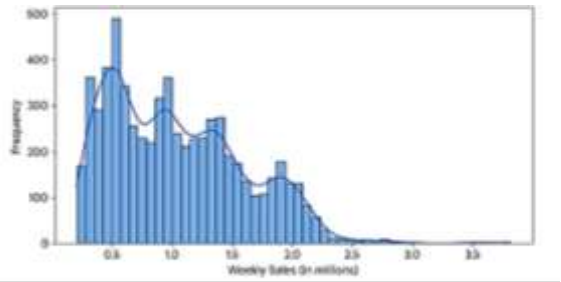

Fig. 2. Kernel Density Estimate for weekly sales

Figure 2 depicts the Weekly Sales (in millions) distribution, dark blue line identifying the trend of sales. Weekly Sales between 0.0 and more than 3.5 million are put on the x-axis, whereas the frequencies of the occurrences are plotted on the y-axis. The distribution is tilted to the right, meaning that most weekly sales fall into lower ranges. There are fewer instances of high sales, suggesting that they were sporadically boosted, most likely as a result of promotions or special events. KDE curve eliminates these variations, and indicates that weekly sales are not distributed in the same way but depend on several factors, including seasonal patterns, promotional, and market differences in the region, which is good information in understanding and forecasting sales performance.

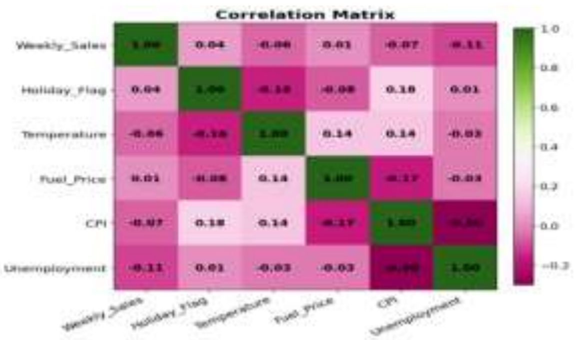

Fig. 3. Plot the Correlation Matrix of Features Of Data

In Figure 3, the correlation of data attributes and the linear relationships of variables is demonstrated. Positive correlations are strong, and are placed along the diagonal, and around the groups of features which are strong, and the weak or negative relationships are placed across the off-diagonal features. It is possible to find redundant predictors, potential multicollinearity and data structure, and in general, do it in a short amount of time using the visualization, therefore, informing both the feature selection and the model refinement.

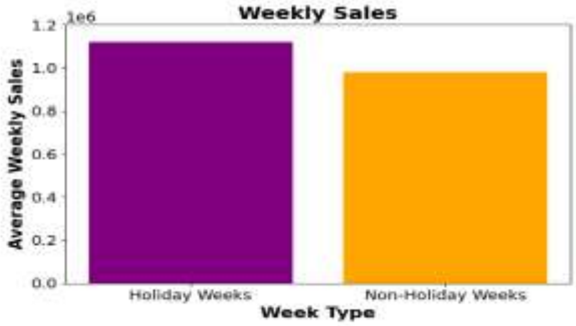

Fig. 4. Bar Graph for Comparison of weekly sales

The relative analysis of the Average Weekly Sales also presented meaningful information on the pattern of consumer spending as Figure 4 illustrates. The average weekly sales in the Holiday Weeks are significantly higher, which points to a considerable increase in the level of purchasing activity during the festive or any special day. During Non-Holiday Weeks, the sales are lower and they average approximately 1,100,000, which means that the consumer behaviour during non-holiday weeks is stable but small. The distinct distinction between the two bars visually highlights the fact that the periods of the holiday seasons are likely to drive higher consumer demand perhaps as a result of promotional discounts, holiday shopping and higher household expenditures. Generally speaking, the visualization greatly indicates that during the seasons of holidays, there is a positive effect on the weekly sales performance, which can be quantified, and the business may use this information to design marketing and inventory strategies.

B. Data Pre-processing

Data pre-processing is essential to producing ML models that are correct. Several procedures have been used to guarantee that datasets are clean and appropriate for analysis.

- Handle missing values: To maintain data integrity, rows in store_data with missing values are eliminated. Only the desired numeric columns are included in a new DataFrame called data_numeric.
- Date time conversion: To enable time-based analysis, the 'Date' column is sorted, transformed to datetime format, and assigned as the index.

C. Outlier Detection using Z-score

Z-scores are then used to identify and eliminate outliers. To avoid extreme values influencing the analysis, data points with absolute Z-scores larger than 2.5 are filtered away. Equation (1) provides the z-score approach formula.

$$x_i = \frac{x-\mu_f}{\sigma_f} \tag{1}$$

Where μ_f is the feature vector f mean, σ_f is its standard deviation, and x_i is the normalized value of x in feature vector f.

D. Feature Engineering

The feature engineering process means the improvement and conversion of the data into a better format to be used in its analysis and modelling. It starts with the transformation of the Date column into a datetime clause and retrieval of time-related elements of the year, month, week number, and day of the week. The lag features of Weekly Sales are then generated by shifting the series by one or three periods, and the missing values are filled with a 0 after which a four period rolling mean is performed to capture trends without data leakage. The Is Holiday column is changed to an integer type and the original Date column is eliminated because the information has already been used.

E. Data Splitting

The pre-processed dataset was initially divided into training and test sets at an 80% to 20% split. This division allows the model to be trained on the larger 80% portion and then independently evaluated on the remaining 20% to assess its generalization performance and accuracy on unseen data.

F. Propose XGBoost model

A number of steps make up the XGBoost architecture, including feature processing, boosting iterations, and assessment [22]. An initial prediction, often set to 0, is the first step in the boosting method. In the first iteration, a decision tree is fitted to represent the negative gradients (errors), and the gradient of the loss is computed in relation to the current predictions [23]. The result is then used to update this tree to provide model predictions. The prediction is generated when each new tree is added in the subsequent rounds, which add trees in a sequential fashion to correct the residual mistakes of the earlier iterations. By adjusting to the residual errors and learning to minimize them, each weak learner or tree contributes to the adjustment of the model predictions. Regularization and shrinkage are critical in the structure such that the complexity of each tree is regulated to avoid overfitting by punishing deeper or more complicated trees [24]. The mathematical formulation of the XGBoost algorithm is based on an ensemble of decision trees, where it incrementally builds trees to minimize a loss function. The objective function of XGBoost consists of two parts: a loss function and a regularization term, as shown in Equation (2).

$$\mathcal{L}(\theta) = \sum_{i=0}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{k} \Omega(f_k) \tag{2}$$

Where $y_i, \hat{y}_i$ is a loss function that quantifies the discrepancy between the true label y_i and the prediction $\hat{y}_i$. Regression and classification often employ the squared error and logistic loss, respectively. To avoid overfitting, a regularization term called $\Omega(f_k)$ penalizes the trees' complexity. k is the ensemble's tree count. The kth tree in the model, or f_k, is a decision tree.

G. Performance Metrics

An essential step in determining evaluating the process's quality involves assessing the outcomes. The results of the computations evaluated using the specified parameters:

1) R-Squared (R^2)

A regression line's fit to dispersed data is assessed using the R^2 score. For comparable datasets, higher R^2 values indicate a smaller discrepancy between the expected and actual data. On a scale of 0 to 1, it calculates the correlation between the projected and actual data. Equation (3) provides it:

$$R^2 = 1 - \frac{SSR}{TSS} \quad (3)$$

2) Mean Absolute Error (MAE)

It is the mean absolute discrepancy between the test sample's actual observations and expectations. The definition of MAE is as follows Equation (4):

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |Y_i - \hat{Y}_1| \quad (4)$$

3) Root Mean Squared Error (RMSE)

The square root of the mean of squares of all the mistakes is the Root Mean Squared error, also known as the Root Mean Squared Deviation [25]. Once more, RMSE indicates the proximity of the line of best fit to the collection of points shown in Equation (5).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (Y_i - \hat{Y}_1)^2} \quad (5)$$

These evaluation metrics are utilized to evaluate and compare the effectiveness of different classification models in accurately identifying various types of performance.

IV. Result Analysis and Discussion

The experiments were conducted in a high-performance computing environment using a 2X-large virtual machine equipped with 8 cores, 64 GB RAM, and 40 GB of disk space. The development platform was Jupyter Notebook which is accessed via Anaconda Navigator. It was implemented in Python and with such libraries as TensorFlow, Keras, Scikit-learn, Pandas, NumPy, Seaborn, Matplotlib, and Imbalanced-learn. The XGBoost model has had a good predictive performance, as shown in the results in Table II, with an R^2 of 0.99% showing that the predictors are strongly correlated with the target variable. The above findings indicate that the model has a greater capacity for making correct, high-reliability predictions and is thus active in real-world predictive systems.

TABLE II. EVALUATION METRICS OF PROPOSED MPDEL FOR SEASONAL SALES PREDICTION ON WALMART DATA

Metrics	XG BOOST
R-Square	0.99
MAE	1226.47
RMSE	1700.98

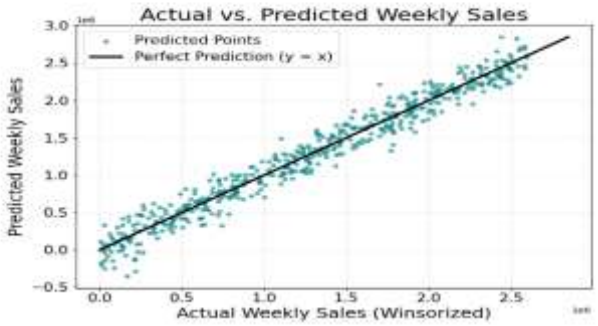

Fig. 5. Actual vs. Predicted Weekly Sales (XGBoost Model)

The high predictive ability of the XGBoost algorithm in the prediction of sales at a weekly level is shown in Figure 5. The x-axis is the Actual Weekly Sales (Winsor zed), whereas the y-axis is the Predicted Weekly Sales. Every blue dot is a pair of real and predicted values which is tightly clustered along the red dotted line, which is Perfect Prediction (y = x) and is almost linear. Close proximity of these data points to the ideal line is evidence of the outstanding predictive power of the model.

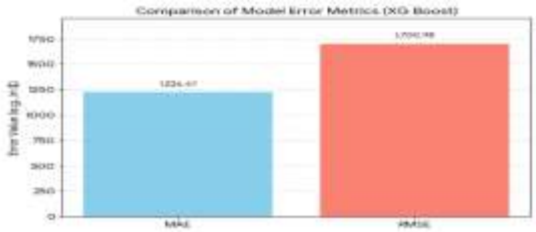

Fig. 6. Bar Graph for Comparision of Error Matrix

Figure 6 compares the values of two key regression metrics: MAE and RMSE. The y-axis represents the Error Value (e.g., in). The chart shows that the MAE (represented by the light blue bar) is 1,226.47, while the RMSE (represented by the salmon-colored bar) is 1,700.98 . The difference between the two bars is notable, with the RMSE being significantly higher than the MAE. This result suggests that while the average error (MAE) is relatively low, the model likely made some larger individual errors or outliers, as the squaring operation in the RMSE formula penalizes them more heavily.

A. Comparative Analysis

This section presents a comparative study of ML models used on big datasets for a challenging predictive regression task. Among the evaluated models, the XGBoost classifier demonstrated exceptional performance, achieving the highest (R^2) value at 99.99 % as present in Table III. This indicates its outstanding capability to explain nearly all the variance in the target variable. DT had a close second with a powerful (R^2) of 93.75% which indicates high generalization strength. These findings make XGBoost the best model to use in this predictive task in this research as it provides close to a perfect predictive accuracy in this metric.

TABLE III. PERFORMANCE COMPARISON OF MACHINE LEARNING MODELS FOR SEASONAL SALES PREDICTION

Method	R2	MAE
GB [26]	81	26.11
DT[27]	96.96	77,388.08
RF[28]	93.75	1937.81
XGBoost	99.99	1226.47

The significance of this research is that it has proposed a highly viable and practical remedy to the seasonal sales forecasts, which is highly essential in effective inventory planning, resource distribution and promotion decisions making in large retail outlets. The study based on the Walmart Sales Forecasting dataset demonstrates that it is possible to identify meaningful trends with the help of ML that are determined by holidays, seasonality, and per-store. The paper is novel since it combines EDA-based feature engineering with advanced temporal characteristics such as lag values and rolling statistics, which enhance the quality of the model to comprehend the dynamics of weekly sales. In addition, the application of an optimized XGBoost regression model is a better and more powerful option compared to the conventional predictive techniques that are effective in modeling the nonlinear relationships and reducing the prediction errors. The excellent results obtained justify the novelty of the suggested framework and outline its prospects of implementing it in the real world in retail forecasting systems.

V. Conclusion and Future Scope

The use of sophisticated ML models for sales forecasting has grown in importance in the quickly changing field of retail analytics. For inventory management, marketing, customer service, and business financial planning in the retail sector, sales forecasting is the most difficult task. This paper introduces a powerful ML model to predict seasonal sales with the help of the Walmart Sales Forecasting dataset. The proposed method resulted in the highly accurate forecasting performance due to extensive pre-processing, feature engineering based on EDA, and the utilization of an optimized XGBoost regression model, where the R^2 was 0.99, and the predicted errors were low. These findings show that the model has a great potential of depicting complex sales trends that are affected by holidays and seasonal trends. The study is however limited. The model mainly emphasizes on numbers and time without considering any other possible external variables like weather, economic variables and competitor activity, which might also affect sales. Also, it can be noted that the Winsorization of extreme values can reduce the sensitivity of the model to occasional but influential sales spikes. This study can be extended by future research to include more contextual features and use DL models like LSTMs or Temporal Fusion Transformers, and apply the framework to multi-store or multi-region forecasting. This will contribute to the generalization and the applicability in the real-life situations.

REFERENCES

[1] J. Thomas, K. V. Vedi, and S. Gupta, "Enhancing Supply Chain Resilience Through Cloud-Based SCM and Advanced Machine Learning: A Case Study of Logistics," *J. Emerg. Technol. Innov. Res.*, vol. 8, no. 9, 2021.

[2] A. Parupalli, "The Evolution of Financial Decision Support Systems : From BI Dashboards to Predictive Analytics," *KOS J. Bus. Manag.*, vol. 1, no. 1, pp. 1–8, 2023.

[3] K. M. R. Seetharaman and S. Pandya, "Importance Of Artificial Intelligence In Transforming Sales, Procurement, And Supply Chain Processes," *Int. J. Recent Technol. Sci. Manag.*, vol. 8, no. July, pp. 140–148, 2023.

[4] U. Vikas, K. Sunil, R. S. Hallikar, P. Deeksha, and R. Kumar P, "A Comprehensive Study on Demand Forecasting Methods and Algorithms for Retail Industries," *J. Univ. Shanghai Sci. Technol.*, vol. 23, no. 06, pp. 409–420, Jun. 2021, doi: 10.51201/JUSST/21/05283.

[5] V. Prajapati, "Enhancing Supply Chain Resilience through Machine Learning- Based Predictive Analytics for Demand Forecasting," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 11, no. 3, 2025.

[6] K. M. R. Seetharaman, "Digital Transformation in Retail Sales: Analyzing the Impact of Omni-Channel Strategies on Customer Engagement," *J. Glob. Res. Math. Arch.*, vol. 10, no. 12, 2023, doi: 10.5281/zenodo.15280578.

[7] K. M. R. Seetharaman and S. Pandya, "Leveraging Ai And Iot Technologies For Demand Forecasting In Modern Supply Chain," *Int. J. Recent Technol. Sci. Manag.*, vol. 9, no. 6, 2024.

[8] V. Kumar and M. L., "Predictive Analytics: A Review of Trends and Techniques," *Int. J. Comput. Appl.*, vol. 182, no. 1, pp. 31–37, Jul. 2018, doi: 10.5120/ijca2018917434.

[9] A. Borucka, "Seasonal Methods of Demand Forecasting in the Supply Chain as Support for the Company's Sustainable Growth," *Sustainability*, vol. 15, no. 9, p. 7399, Apr. 2023, doi: 10.3390/su15097399.

[10] H. Kali and G. Modalavalasa, "Artificial Intelligence (AI)-Driven Business Intelligence for Enhancing Retail Performance with Customer Insights," *Asian J. Comput. Sci. Eng.*, vol. 9, no. 4, 2024, doi: 10.22377/ajcse.v10i2.210.

[11] C. Patel, "A Survey of Data-Driven Customer Segmentation Methods for Targeted Marketing Campaigns," vol. 3, no. 3, pp. 154–162, 2023, doi: 10.56472/25832646/JETA-V3I7P119.

[12] Q. Li and M. Yu, "Achieving Sales Forecasting with Higher Accuracy and Efficiency: A New Model Based on Modified Transformer," *J. Theor. Appl. Electron. Commer. Res.*, vol. 18, no. 4, pp. 1990–2006, Nov. 2023, doi: 10.3390/jtaer18040100.

[13] S. P. Kalava, "Revolutionizing Customer Experience: How CRM Digital Transformation Shapes Business," *Eur. J. Adv. Eng. Technol.*, p. 4, 2024.

[14] R. Q. Majumder, "Machine Learning for Predictive Analytics : Trends and Future Directions," *Int. J. Innov. Sci. Res. Technol.*, vol. 10, no. 4, 2025.

[15] A. R. Bilipelli, "End-to-End Predictive Analytics Pipeline of Sales Forecasting in Python for Business Decision Support Systems," *Int. J. Curr. Eng. Technol.*, vol. 12, no. 6, pp. 819–827, 2022.

[16] M. P. Rahman Mahin, M. Shahriar, R. R. Das, A. Roy, and A. W. Reza, "Enhancing Sustainable Supply Chain Forecasting Using Machine Learning for Sales Prediction," *Procedia Comput. Sci.*, vol. 252, pp. 470–479, 2025, doi: 10.1016/j.procs.2025.01.006.

[17] C. Deng *et al.*, "Prediction of retail commodity hot-spots: a machine learning approach," *Data Sci. Manag.*, vol. 8, no. 4, pp. 414–422, Dec. 2025, doi: 10.1016/j.dsm.2025.02.003.

[18] S. Balaji and D. Manikavelan, "Stacking Based Enhanced Sales Forecasting for E-commerce Using Ensemble Learning approaches with Multi-Scaling Data," in *2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL)*, IEEE, Feb. 2025, pp. 221–228. doi: 10.1109/ICSADL65848.2025.10933090.

[19] D. S. AbdElminaam, M. Mohamed, S. Khaled, F. Hany, M. Magdy, and Y. Sherif, "Leveraging Machine Learning for Accurate Store Sales Prediction: A Comparative Study," in *2024 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, IEEE, Nov. 2024, pp. 355–362. doi: 10.1109/MIUCC62295.2024.10783509.

[20] S. Sharif, M. T. Tamang, A. Nepal, and W. Elmedany, "A Comparative Study of Sales Prediction Using Machine Learning Models: Integration of PySpark and Power BI," in *2024 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, IEEE, Nov. 2024, pp. 137–142. doi: 10.1109/3ict64318.2024.10824620.

[21] H. Upadhyay, S. Shekhar, A. Vidyarthi, R. Prakash, and R. Gowri, "Sales Prediction in the Retail Industry Using Machine Learning: A Case Study of BigMart," in *2023 International Conference on Electrical, Electronics, Communication and Computers (ELEXCOM)*, IEEE, Aug. 2023, pp. 1–6. doi: 10.1109/ELEXCOM58812.2023.10370313.

[22] M. R. R. Deva and N. Jain, "Utilizing Azure Automated Machine Learning and XGBoost for Predicting Cloud Resource Utilization in Enterprise Environments," in *2025 International Conference on Networks and Cryptology (NETCRYPT)*, IEEE, May 2025, pp. 535–540. doi: 10.1109/NETCRYPT65877.2025.11102235.

[23] N. Prajapati, "The Role of Machine Learning in Big Data Analytics: Tools, Techniques, and Applications," *ESP J. Eng. Technol. Adv.*, vol. 5, no. 2, 2025, doi: 10.56472/25832646/JETA-V5I2P103.

[24] I. Hussain, K. B. Ching, C. Uttraphan, K. G. Tay, I. Memon, and S. A. Memon, "Predicting Monthly Wind Speeds Using XGBoost: A Case Study for Renewable Energy Optimization," *Processes*, vol. 13, no. 6, p. 1763, Jun. 2025, doi: 10.3390/pr13061763.

[25] A. V Tatachar, "Comparative Assessment of Regression Models Based On Model Evaluation Metrics," *Int. Res. J. Eng. Technol.*, vol. 8, no. 9, pp. 853–860, 2021.

[26] P. K. Yadav, V. Kumar, R. Bhushan, and P. K. Singh, "Analysis of Machine Learning Model for Predicting Sales Forecasting," in *2023 1st International Conference on Advances in Electrical, Electronics and Computational Intelligence, ICAEECI 2023*, 2023. doi: 10.1109/ICAEECI58247.2023.10370914.

[27] C. N. C. *et al.*, "Advancing Retail Predictions: Integrating Diverse Machine Learning Models for Accurate Walmart Sales Forecasting," *Asian J. Probab. Stat.*, vol. 26, no. 7, pp. 1–23, 2024, doi: 10.9734/ajpas/2024/v26i7626.

[28] B. Yao, "Walmart Sales Prediction Based on Decision Tree, Random Forest, and K Neighbors Regressor," *Highlights Business, Econ. Manag.*, vol. 5, pp. 330–335, Feb. 2023, doi: 10.54097/hbem.v5i.5100.

