

# An Effective Predicting E-Commerce Sales & Management System Based on Machine Learning Methods

Manikanth Sarisa <sup>1,\*</sup>, Venkata Nagesh Boddapati <sup>2</sup>, Gagan Kumar Patra <sup>3</sup>, Chandrababu Kuraku <sup>4</sup>, Siddharth Konkimalla <sup>5</sup>, Shravan Kumar Rajaram <sup>6</sup>

<sup>1</sup> Sr Application Developer, Bank of America, USA

<sup>2</sup> Support Escalation Engineer, Microsoft, USA

<sup>3</sup> Senior Solution Architect, Tata Consultancy Services, India

<sup>4</sup> Senior Solution Architect, Mitaja Corporation, USA

<sup>5</sup> Network Development Engineer, Amazon Com LLC, USA

<sup>6</sup> Network Engineer, AT & T, USA

\*Correspondence: Manikanth Sarisa (Mk2703@outlook.com)

**Abstract:** Due to influence of Internet, this e-commerce sector has developed rapidly. Most of the online retailing or selling businesses are seeking for way for predicting their products demand. Sales forecasting may help retailers develop a sales strategy that will enhance sales and attract more money and investment. The current research work puts forward a machine learning framework to forecast E-commerce sales for strategic management using a dataset of E-commerce transactions. With 70 percent of the data for train and 30 percent for test, three models were produced, namely, Random Forest, Decision Tree, and XGBoost. In order to evaluate the models, performance measures inclusive of R-squared ( $R^2$ ) and Root Mean Squared Error (RMSE) were employed. Thus, the XGBoost model was the most accurate in marketing predictive capabilities for E-commerce sales with the  $R^2$  score of 96.3%. This has demonstrated the increased capability of XGBoost algorithm to forecast E-commerce monthly sales more accurately than other models and can assist decision makers for managing inventory and arriving smart and quick decisions in this rapidly growing E-commerce market. The findings reiterate the importance of using advanced analytics in order to drive effectiveness and customer experience within E-commerce sector.

**Keywords:** Sales, Forecast, Pattern Recognition and Systems ML

## How to cite this paper:

Sarisa, M., Boddapati, V. N., Patra, G. K., Kuraku, C., Konkimalla, S., & Rajaram, S. K. (2020). An Effective Predicting E-Commerce Sales & Management System Based on Machine Learning Methods. *Journal of Artificial Intelligence and Big Data*, 1(1), 75–85. Retrieved from <https://www.scipublications.com/journal/index.php/jaibd/article/view/1110>

**Received:** August 23, 2020

**Revised:** November 20, 2020

**Accepted:** December 22, 2020

**Published:** December 27, 2020



**Copyright:** © 2020 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The faster growing rate and consumption of e-commerce technologies, many consumers prefer to buy on different e-commerce platforms. Users may shop whenever and wherever they please and with convenience as opposed to having to buy items physically from traditional outlets. They also do not have to wait for the weekend before they go shop and shop till they drop. Also, platforms have many different types of products in several designs so the clients can buy what they need without going outside [1]. There are many benefits of online shopping to the customer, though since e-commerce sites are mostly virtual, there are certain flaws with the goods that are retailing on such sites. These issues include inconsistent descriptions of products and their actual quality, subpar after-sales care, and more. Playing particular emphasis on electronic commerce platforms, it is therefore quite very important to undertake a sentiment analysis on the commodity assessment of the acquired items [2].

Big data analytics has become popular in recent years because of the avalanche of data generated from multiple sources, including social media, consumer loyalty programs, and point-of-sale systems. Supermarkets may examine this data and learn about

consumer behaviour, preferences, and trends with the use of big data analytics [3]. Using machine learning algorithms, predictive models that discover patterns, estimate future sales trends and give supermarkets useful insights may be created from the data.

For companies that sell retail goods, forecasting sales is crucial since it helps with inventory control. An essential part of inventory optimisation is forecasting product sales. Because certain e-commerce companies, as everyone is aware, sell their own distinct things online. An e-commerce platform of that kind often has to keep a careful check on its stock. Sales forecasting has several advantages, including budgetary distribution, goal-setting, audience targeting, performance evaluation, and many more. In an E-commerce business, sales forecasting and projection are crucial as they indicate the effectiveness of the sales team, aid in managing the budget and informing decision-making, and establish the target market [4]. A machine learning algorithm may be used to analyse the significant patterns and variables included in this data, which will allow for a very accurate forecast of sales. Based on data, a machine learning model is learnt to look for patterns that repeat themselves in order to make future occurrences predictions. Machine learning models have emerged as indispensable tools in grocery sales analysis. Its simplicity and interpretability make it an attractive choice for modelling sales trends.

#### ***A. Motivation and Contribution of Study***

The swift expansion of e-commerce has resulted in a rise in the intricacy of data, as market circumstances, seasonal patterns, and customer preferences influence the amount of transactions. Accurate sales predictions are vital for E-commerce businesses to manage inventory, enhance customer satisfaction, and optimize resource allocation. However, conventional forecasting methods often fall short in capturing these dynamic trends. This study is motivated by the need to leverage machine learning to create a robust, predictive sales model capable of addressing these challenges, thereby aiding in decision-making and helping businesses adapt to fluctuations in demand. This study contributes to the field of E-commerce analytics by developing a predictive system to forecast sales trends with improved accuracy. The following are the study's primary contributions:

- Collected and analyzed E-commerce transaction data spanning multiple years to identify key sales and trend patterns.
- Applied preprocessing techniques, including missing value handling and outlier removal, to ensure data consistency.
- Utilised label encoding to convert categorical features, enhancing compatibility with machine learning models.
- Selected and refined features based on transaction volume, pricing, and economic factors to improve model accuracy.
- Implemented and optimised models (Random Forest, Decision Tree, and XGBoost) for effective sales prediction.
- Evaluated model performance using metrics such as R-squared and RMSE, ensuring reliable assessment of prediction accuracy and model robustness.

#### ***B. Structure of paper***

The paper's structure is arranged as follows: A overview of earlier studies on the topic of predicting e-commerce sales is given in Section II. The research process is described in Section III, along with the methodology, data pretreatment, and model selection that were employed. In Section IV, the models' performance is analysed and the experimental data is presented. Lastly, Section V addresses the ramifications of the findings and provides a summary of the major conclusions.

## **2. Literature Review**

The several machine learning methods that may be used to predict e-commerce sales and management systems are described in recent papers that are included in this section. Below are a few background studies:

To predict Walmart's sales issue, the following XGBoost sale prediction model is proposed by Niu (2020) which composed of XGBoost algorithm with comprehensive feature engineering processing. The approach utilised in this work could effectively employ one or the other characteristics of several dimensions to generate efficient prediction. Based on the datasets of Walmart store sales provided by the Kaggle competition, this paper evaluates the XGBoost sale forecast model. Having compared with the other machine learning algorithms, experimental data shows that our strategy is better. The RMSSE measure of this work is 0.141 and 0.113 less than the RMSSE of the Ridge and Logistic Regression algorithms. Also, this work aims to determine the importance of the attributes and gives some useful recommendations [5].

In this paper, Wisesa, Adriansyah and Khalaf, (2020) a succinct examination of B2B sales dependability via machine learning methods. A variety of sales forecast techniques and treatments are explained in the research's final section. The data analysis results in the evaluation of the performance of each prediction model provide the best-adapted prediction model to be used on the B2B sales trend projection. The findings of the projection, estimation, and analysis are summed up in terms of reliability and validity of effective forecasting and prediction techniques. It is anticipated that the analysis's output would provide forecasting data that is dependable, precise, and efficient—a crucial tool for projecting sales. Studies have demonstrated that the Gradient Boost Algorithm, with MSE = 24,743,000,000.00 and MAPE: 0.18, has strong forecasting and future B2B sales prediction accuracy [6].

In this paper, Ding et al., (2020) suggested using CatBoosting to create a sales forecasting system. To train the algorithm, the Walmart sales dataset, currently the biggest dataset in this sector, is employed. Feature engineering was applied well in order to improve the model's prediction time and accuracy. Our model achieves an RMSE of 0.605 in the experiments, outperforming conventional machine learning techniques like SVM and linear regression. Our method's capacity to generalise on other bespoke datasets is improved and its potential utility is expanded since it requires less fine-tuning than previous approaches [7].

Kulshrestha and Saini, (2020) in the current paper, analyzed an e-commerce company's selling data set, divided it into different quarters, and calculated the revenue from sales by that quarter. Following that, the final dataset was separated to the one having 70% for training and the other – 30%, for testing. We will analyse the most sold commodities and their frequency of purchase every quarter, as well as project income for the upcoming quarters using a machine learning algorithm. Then, give the company organisation the analysis results and the predicted client purchase patterns so they can develop a plan to maintain and accumulate for their inventory management and get a competitive edge [8].

In this study, Kaneko and Yada, (2016) make a sale prediction model by applying point of sale (POS) data collected for three years from a retail business. The model is built to predict changes that occur in either occurrence or amount of sale on any particular day as compared to the previous day. Therefore, when using a reachable deep learning model that incorporates the L1 regularisation, it became possible for the system to predict sales with accuracy only at 86 percent. The retail store's inventory has been carefully arranged into categories. Increased from tens to thousands of characteristics of the product categories did not result in a failure of the predicted accuracy to be decreased by more than 7%. However, when using logistic regression the accuracy percentage reduced by about 13%. These findings suggest that building models with multi-attribute variables is a great use for deep learning. The current study shows that deep learning is useful for assessing retail shop POS data [9].

The following [Table 1](#) provides the background study comparison between data, performance, and limitations.

**Table 1. Summary of literature review on e-commerce-based sales prediction using machine learning**

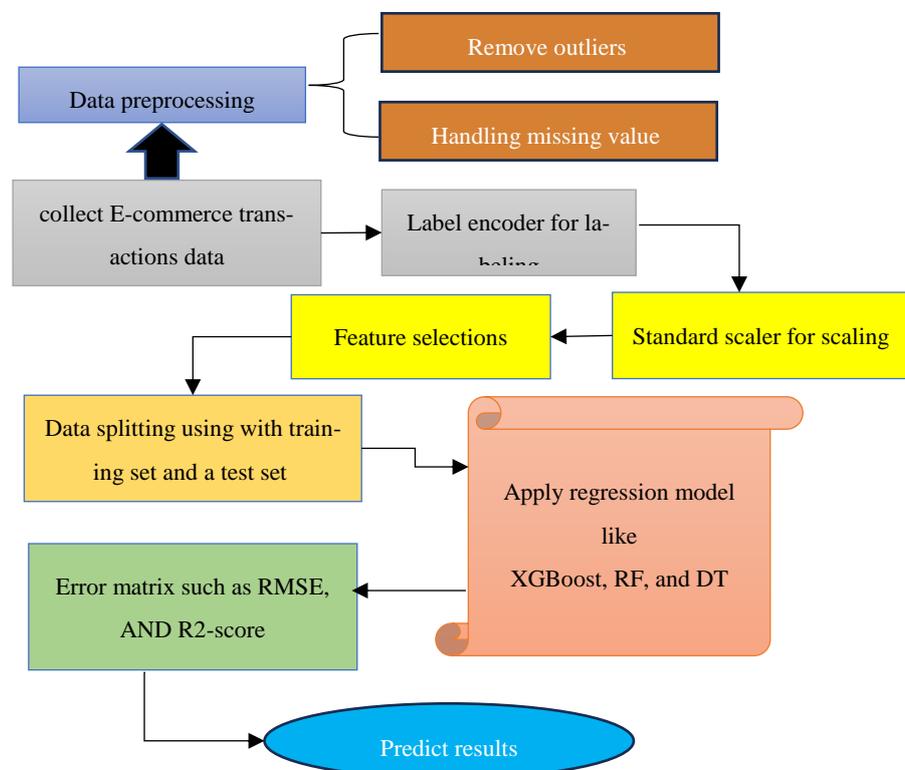
Author	Methodology	Data	Findings	Limitations
Niu (2020)	XGBoost with feature engineering for sales prediction	Walmart sales dataset from Kaggle	XGBoost outperformed Logistic Regression and Ridge Regression; RMSSE was 0.141 and 0.113 lower than these models, respectively. Feature importance was ranked.	Limited comparison to other models; only Walmart sales data used.
Wisesa, Adriansyah & Khalaf (2020)	Gradient Boosting for B2B Sales Prediction	B2B sales data	Gradient Boosting had good accuracy with MSE = 24,743,000,000 and MAPE = 0.18. Generated reliable, accurate sales forecast data.	Lack of detail on specific data attributes; limited evaluation beyond MSE and MAPE.
Ding et al. (2020)	CatBoost with feature engineering for sales prediction	Walmart sales dataset	CatBoost outperformed traditional models such as SVM and linear regression with an RMSE of 0.605. Required less fine-tuning and improved generalisation.	Only Walmart sales data used; limited comparison to other ensemble models.
Kulshrestha & Saini (2020)	Sales prediction based on ML, split data by quarters and predicted future income.	E-commerce sales data	Predicted future income and analyzed frequently sold commodities. Provided strategic insights for inventory and business planning.	Did not specify the exact ML algorithm used; limited comparison to other models.
Kaneko & Yada (2016)	Deep learning model with L1 regularization for sales prediction	3 years of POS data from a retail store	Achieved 86% accuracy for sales prediction. Deep learning model outperformed Logistic Regression, with accuracy falling by 7% versus 13% for Logistic Regression.	No mention of computational complexity or scalability for larger datasets. Only tested on retail POS data.

### A. Research gaps

Current research on sales prediction using machine learning has shown promising results but still presents several gaps. Most studies focus on specific industries like retail and e-commerce, limiting the generalizability of models across diverse sectors. Additionally, while feature engineering enhances prediction accuracy, model interpretability remains a challenge, particularly in complex algorithms like deep learning and ensemble methods. Scalability and computational efficiency for large datasets are also underexplored, with limited attention given to the practicality of these models in big data environments. Furthermore, while some algorithm comparisons exist, there is a lack of research into hybrid models that could leverage the strengths of multiple approaches. Lastly, the reliance on single-source datasets restricts the robustness of the findings, indicating a need for more diverse data to validate the effectiveness of these models in broader contexts.

## 3. Methods and Materials

For Predicting E-Commerce Sales & Management System Based on Machine Learning Methods use E-commerce transaction records collected in China from 2005 to 2019, capturing annual transaction volumes and various influencing factors, including resource availability, transaction activity, and economic development indicators. After data collection, pre-process the data for data cleaning. In pre-processing, including handling missing values, outlier treatment, scaling, and feature selection to improve model accuracy. A 70-30 data split was used for training and testing Random Forest, Decision Tree and XGBoost were the selected machine learning models. In order to improve the system's predicted accuracy and offer useful insights for E-commerce management, evaluation measures like R-squared and RMSE were employed to evaluate The performance of the model. The process of designing the research is illustrated in Figure 1; data flow diagram.



**Figure 1.** Flowchart for Predicting E-Commerce Sales& Management System

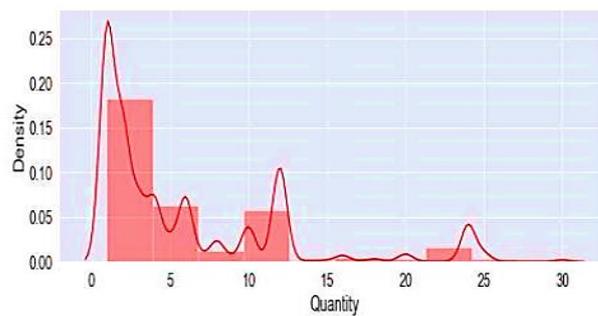
#### A. Data source

The sample set of e-commerce transactions was gathered in China between 2005 and 2019 and included the amount of e-commerce transactions annually as well as a variety of contributing factors. The primary determinants of E-commerce transactions are the level of economic development, transaction level, and fundamental resource availability. The analysis of data with EDA is given in below:



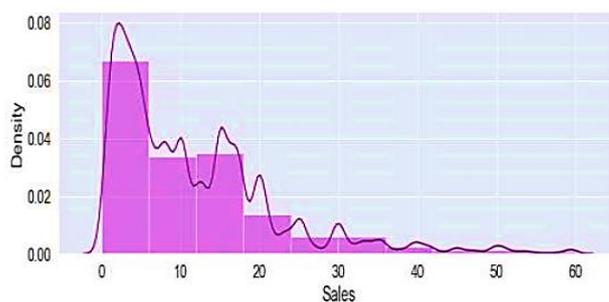
**Figure 2.** Unit Price plot

The density plot of Unit Price in [Figure 2](#) illustrates the distribution of unit prices within the dataset, with prices ranging from 0 to 5 on the horizontal axis and density values on the vertical axis, peaking at 0.5. The fluctuating line curve reflects the frequency of different unit prices, showing where prices are most concentrated. This plot helps visualize how often certain price ranges appear in the data, providing insights into common pricing trends.



**Figure 3.** Quantity plot

[Figure 3](#) Quantity plot shows a red line graph plotting 'Density' on the vertical scale as 'Quantity', and on the horizontal scale, 'Price'. The density values range from 0 to 0.25, and the quantity values range from 0 to 30. The graph has several peaks, with the highest peak occurring just before the quantity of 5.



**Figure 4.** Density plot for Sales

The sale densities are represented by the density plot as evident in the [Figure 4](#) above. On this figure, the x-axis displays Sales in the range of 0 and 60, while the y-axis is Density in the range of 0 and 0.08. The plot shows a peak in sales around 5-10, indicating that this range has the highest frequency of sales occurrences. As the sales values increase, the density gradually decreases with some fluctuations.



Figure 5. Sales over time

Sales over time in Figure 5 shows sales data from December 2010 to November 2011. It highlights a significant drop in sales in January, followed by a gradual increase with some fluctuations throughout the year, peaking in November. This visual representation can be quite useful for analyzing business trends and making informed decisions.

### B. Data preprocessing

The process of removing all noise and outliers from the chosen data is known as data preparation. This entails purging the facts that include a significant quantity of extra meaning, which indicates less and is not necessary. The key preprocessing steps are present in below:

- **Handle missing value:** To maintain consistency in the data, either apply the mean or median imputation or substitute the average value for the missing variable.
- **Remove Outliers:** For extreme outliers, consider capping or removing values based on business logic. So, all outliers are removed from the dataset.

### C. Label encoding

Label encoding is the process of converting categorical data into numerical form in the domains of the machine learning and data analysis. When working with techniques that require numerical input, it is very useful because most machine learning models can only run with numerical data.

### D. Standard scaler for scaling

Standardisation, another name for the standard scaler, is another well-liked feature scaling method in machine learning. Every feature is converted by the approach to a mean at zero and a unit variance. Although the majority of the data will lie somewhat close to 0, this approach does not restrict the data into a single interval or change its dispersion. This indicates that even after scaling, outliers can still be found in the data. Equation 1 provides a definition of standard scaling.

$$x_{scaled} = \frac{x - \bar{x}}{\sigma} \quad (1)$$

where:  $x_{scaled}$  = a scaled sample point

- $x$  = example point
- $\bar{x}$  = average of the practice examples
- $\sigma$  = the training samples' standard deviation

**Feature selection:** Feature selection basic process for enhancing accuracy of a dataset where data reduction involves the removal of unnecessary information using an assessment index. Therefore, it is crucial to recognise and pick the most pertinent elements from the data and eliminate the ones that are unnecessary or unimportant.

### E. Data splitting

This led to partitioning of the dataset into test and training set. Of the whole data, about 70% was set aside for training, while the remaining 30% was set aside for testing.

#### F. Model Building

The process of modelling involves determining the algorithms that will be applied to the project's study. Several algorithms, including Random Forest, decision trees, and XGBoost, were utilised in this study [10].

##### 1) Random Forest (RF)

Random Forests are one of the ensemble learning technique used in solving regression and classification problems. It constructs many decision trees with randomly selected subsets from the input data and characteristics and then computes an average or performs voting on the future forecasts of all the decision trees. In general, the algorithm is good and versatile; the results obtained are accurate while the output is easily understandable.

##### 2) Decision Tree (DT)

The term that is used when doing machine learning to predict regression and classification is called a decision tree. It constructs a tree like decision model through a repetitive process of binary splitting of the data by adopted variables until a stopping criterion is met. Each branch means a possible end result is present, and each node means a decision based on a feature is present. Decision trees are easy to use an effective means of analyzing numerical and categorical data.

##### 3) XGBoost

It is an ensemble learning system and it operates its predictions by decision trees, called XGBoost. In this respect, it may be applied to regression scenarios by reducing a loss function that measures the difference between the planned and realised target. The mathematical model used in XGBoost regression can be described as Equation (2):

$$y = f(x) \quad (2)$$

where  $y$  is the predicted property price, The input features denoted as vector  $x$  have the features as the number of bedrooms, area in square feet among others. The boost algorithm used here is the XGBoost model,  $f(x)$ .helper $x$  and  $f(x)$  are two terms used to denote the forecast of  $y$ . XGBoost constructs decision trees, the primary parameter needed to optimize is the MSE loss for determination of  $f(x)$ . After the data has gone through ten passes of the above generalization and specialization process, the system produces a final output of an X decision tree forecast. In this paper, I will also show how to use manually the formula of an XGBoost regression model which, in general, has the following appearance (3):

$$y = \sum(k = 1 \text{ to } K) f_k(x) \quad (3)$$

The estimated version of the NDF is expressed as;  $f(x) = \sum(f_k(x)/K)$  – The sum of the forecast output of all the decision trees in the ensemble where  $K$  depicts the overall number of the decision trees in the ensemble, and  $f_k(x)$  represents the forecast of the particular/fiduciary  $k$ -th decision tree. Specifically, the forecast of each tree is the aggregation of the values of its leaves which are obtained during training and weighted. The sum of the decision trees' estimates of the XGBoost model for a specific example of an input ( $x$ ).

#### G. Performance matrix

A key component of developing machine learning projects is model assessment, which facilitates the understanding of model performance and facilitates the explanation

and presentation of model output. The objective then becomes to demonstrate how near the projected values are to the real values because it might be challenging to anticipate a regression model's precise value. Using performance assessment indicators, this study assessed the models.

### 1) R-Squared

The regression model's fit to the data is gauged by its R2 value. A higher value of or R2 show better fit of the model to the empirical findings. R2 values range from 0 to 1. When the R2 number is 1, it implies that the model accurately predicts the response data, whereas an R2 value of 0 mean that the model has no capacity of explaining the variation of the response data around their mean. R2 may be calculated using formula (4):

$$R^2 = \frac{\sum_{i=1}^n (y_i - y_i^p)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

### 2) RMSE (Root Mean Squared Error)

Mean squared error is prevalent, and the standard measure of error is the square root of mean squared error and thus it is. The RMSE reveals how much the points predicted there by use of the model deviate from the actualities. Furthermore, and in terms of the model accuracy, there is evidence of the lower RMSE for both tests. The RMSE calculation formula is (5):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i^p)^2} \quad (5)$$

When taken as a whole, these metrics offer information on how well the model predicts the target variable and how accurate it is.

## 4. Result Analysis and Discussion

The objective of this study was to Predict E-Commerce Sales & Management Systems. In this study used various regression techniques. The following Table 2 provides the XGBoost model performance across the performance.

Table 2. XGBoost model performance on E-Commerce Sales & Management System

Matrix	XGBoost
R2	96.3
RMSE	7.7

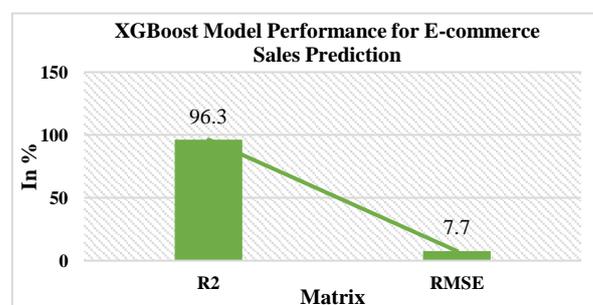
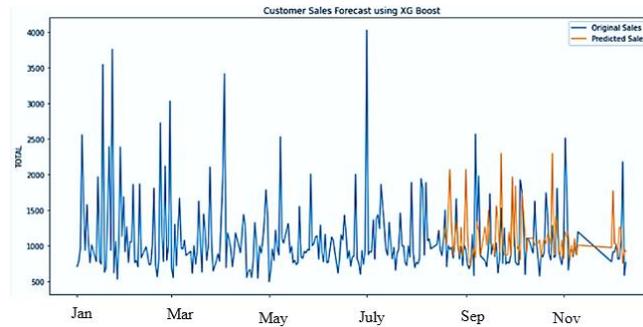


Figure 6. XGBoost model performance

The performance of the XGBoost model shows in above Table 2 and Figure 6. The XGBoost model demonstrates strong performance in sales prediction, achieving an R2 value of 96.3, which indicates excellent variance explanation, and a low RMSE of 7.7,

reflecting accurate predictions close to real auctions values. These indicators demonstrate how well the model performs in providing accurate e-commerce sales projections.



**Figure 7.** Prediction using the XGBoost model

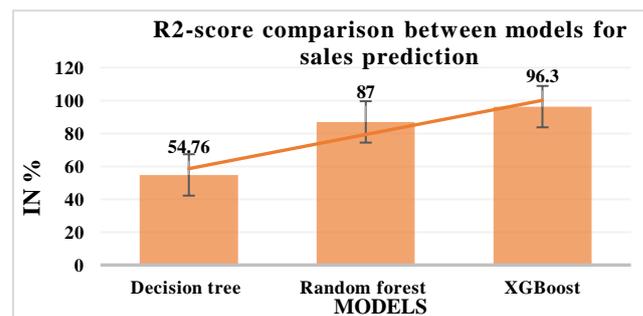
The graph in [Figure 7](#) illustrates the Customer Sales Forecast using the XGBoost model. It compares the actual sales data (in blue) with the predicted sales (in orange) from January to November. The predicted sales closely follow the pattern of the actual sales, indicating that the XGBoost model is performing well in forecasting sales trends.

**A. Comparative Analysis**

This work entails a performance matrix-based comparative analysis of the chosen machine learning models for sales prediction. This comparison, is based on the transaction dataset and utilized the RF [11], DT [12], and xgboost model.

**Table 3. Model Comparison for sales prediction on transaction data**

Models	R2
Decision tree	54.76
Random forest	87
XGBoost	96.3



**Figure 8.** Bar graph for comparison of the model’s performance

[Figure 8](#) shows the comparison of sales prediction models reveals significant differences in performance based on their R<sup>2</sup> values. Analysing the results, we found that For the DT model, R<sup>2</sup> is 54.76, and for the RF model, it is 87. Nonetheless, the XGBoost model has an outstanding prediction accuracy with the high R<sup>2</sup> of 96.3, which mean that among three models presented, it will be the most efficient in sales forecasting.

**5. Conclusion and Future Scope**

The turnover of retail goods and the share of Internet trading has grown rapidly over the past two to three years. Consequently, as consumers employ more internet services in making decision about buying options between actual general physical stores it appears

that the present research rightly demonstrated the possibility of using machine learning for predicting E-commerce sales. The aforementioned analysis employed transaction dataset. Three, this paper deals with the application of machine learning approaches to sales forecasts of E-commerce using the XGBoost model. XGBoost model also has a high value of R-squared of 96.3% which again assists in the accurate prediction of the dataset over both the Random Forest and Decision Tree models. With this improvement in mind, it opens a number of opportunities for E-Commerce management for managing inventories and for making sound decisions by the business. Nevertheless, as a contribution towards development of the rich and improved prediction model such additional data in the future can be combined with this of kind, such as the tweets, for instance, the trends on Twitter or the consumer sentiment. Thus, work in the area of new solutions using deep learning or in the focus on integrating various algorithms would also enhance the results obtained above. However, increasing the extent of data, adding more geographical and time data may provide a wider perspective of E-commerce sale characteristics which in turn may assist in designing better stable and elastic sales predicting structure. To overcome these limitations and further experiences and research avenues that will be important for E-commerce analytics, the following are implied.

## References

- [1] R. Liang and J. Qiang Wang, "A Linguistic Intuitionistic Cloud Decision Support Model with Sentiment Analysis for Product Selection in E-commerce," *Int. J. Fuzzy Syst.*, 2019, doi: 10.1007/s40815-019-00606-0.
- [2] P. Ji, H. Y. Zhang, and J. Q. Wang, "A Fuzzy Decision Support Model with Sentiment Analysis for Items Comparison in e-Commerce: The Case Study of <http://PConline.com>," *IEEE Trans. Syst. Man, Cybern. Syst.*, 2019, doi: 10.1109/TSMC.2018.2875163.
- [3] S. K. R. A. Sai Charan Reddy Vennapusa, Takudzwa Fadziso, Dipakkumar Kanubhai Sachani, Vamsi Krishna Yarlagadda, "Cryptocurrency-Based Loyalty Programs for Enhanced Customer Engagement," *Technol. Manag. Rev.*, vol. 3, no. 1, pp. 46–62, 2018.
- [4] J. Li, T. Wang, Zhengshi, and C. Luo, "Machine Learning Algorithm Generated Sales Prediction for Inventory Optimization in Cross-border E-Commerce," *Int. J. Front. Eng. Technol.*, 2019.
- [5] Liu, G., Nguyen, T. T., Zhao, G., Zha, W., Yang, J., Cao, J., ... & Chen, W. (2016, August). Repeat buyer prediction for e-commerce. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 155-164).
- [6] Jia, R., Li, R., Yu, M., & Wang, S. (2017, July). E-commerce purchase prediction approach by user behavior data. In *2017 international conference on computer, information and telecommunication systems (CITS)* (pp. 1-5). IEEE.
- [7] Cirqueira, D., Hofer, M., Nedbal, D., Helfert, M., & Bezbradica, M. (2019, September). Customer purchase behavior prediction in e-commerce: A conceptual framework and research agenda. In *International workshop on new frontiers in mining complex patterns* (pp. 119-136). Cham: Springer International Publishing.
- [8] Cen, Y., Zhang, J., Wang, G., Qian, Y., Meng, C., Dai, Z., ... & Tang, J. (2019). Trust relationship prediction in alibaba E-commerce platform. *IEEE Transactions on Knowledge and Data Engineering*, 32(5), 1024-1035.
- [9] Y. Kaneko and K. Yada, "A Deep Learning Approach for the Prediction of Retail Store Sales," in *IEEE International Conference on Data Mining Workshops, ICDMW*, 2016. doi: 10.1109/ICDMW.2016.0082.
- [10] N. Agarwal, S. Gupta, and S. Gupta, "A comparative study on discrete wavelet transform with different methods," in *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*, IEEE, Mar. 2016, pp. 1–6. doi: 10.1109/CDAN.2016.7570878.
- [11] Rachid, A. D., Abdellah, A., Belaid, B., & Rachid, L. (2018). Clustering prediction techniques in defining and predicting customers defection: The case of e-commerce context. *International Journal of Electrical and Computer Engineering*, 8(4), 2367.
- [12] Qiu, J., Lin, Z., & Li, Y. (2015). Predicting customer purchase behavior in the e-commerce context. *Electronic commerce research*, 15, 427-452.