Enhancing Predictive Models in E-Commerce: A Comparative Study Using XGBoost Across Diverse Scenarios

Jianhao Zhang

Beijing Royal School No.11 Wangfu Street, North Qijia Town, Changping District, Beijing, 102207, China

Abstract. With the growth of the internet, online shopping has become increasingly popular. However, sudden demand spikes during holidays or special events can disrupt market equilibrium, causing stock shortages and logistical challenges. To address these sudden surges in demand, this study utilizes existing online sales data, transforming it into actionable insights. Our strategy involves continuously feeding historical data into selected models to predict future sales volumes. By identifying patterns in the data, we aim to make the predictions more tangible and assess the validity of our approach through a statistical linear regression model. We employed three different models—Gradient Boosting Decision Tree (GBDT),Light Gradient Boosting MachineLight (GBM),and Extreme Gradient Boosting (XGBoost)—to determine which one is the most efficient. After comparing the performance of these models, the results indicate that XGBoost is the optimal choice. The findings suggest that accurate sales predictions enable e-commerce platforms to increase inventory during peak periods, maximize the utility of goods, ensure customer satisfaction, and stimulate transaction activity. This study underscores the importance of accurately forecasting sales and revenue in e-commerce, helping platforms to stay ahead of demand, optimize resource allocation, and maintain market competitiveess.

1 Introduction

With the rapid rise of e-commerce platforms, accurately predicting the total number of commodities sold in the global marketplace has become increasingly critical. These platforms offer immense convenience to consumers, but this convenience also comes with challenges, particularly during special occasions such as holidays when sudden surges in demand for specific products can occur. Such unexpected spikes in demand can disrupt market equilibrium, leading to issues like stock shortages and logistical bottlenecks. Therefore, the ability to accurately forecast sales and revenue is crucial for e-commerce giants like Amazon, as it enables them to anticipate market shifts, optimize resource allocation, and ensure that they can meet consumer demands efficiently.

Corresponding author: YFSLZYS@163.com

Existing studies, such as those by Kulshrestha and Saini on the application of machine learning algorithms, Huo's work on deep learning techniques for sales forecasting, Zhu's research on hybrid models incorporating sentiment analysis, and Zhan et al.'s exploration of ensemble learning methods, provide valuable insights into market prediction methodologies. However, these studies often fall short of addressing the full complexity of accurately forecasting sales in dynamic and rapidly changing market environments. Despite the valuable insights from these studies, there remains a gap in effectively forecasting sales within such volatile market conditions.

Our research aims to bridge this gap by utilizing historical sales data to train advanced predictive models. The primary goal of our study is to predict the monthly sales of commodities for the upcoming year by analyzing historical sales data. Through this predictive model, our research seeks to provide actionable insights that can significantly improve the management of market dynamics. By anticipating fluctuations in consumer demand and adjusting inventory levels accordingly, e-commerce platforms can better navigate the challenges of market volatility, maintain a stable supply chain, enhance customer satisfaction, and support the continued growth and sustainability of the e-commerce sector.

In our approach, we opted to use XGBoost due to its excellent performance in handling large-scale data and complex features, as well as its ability to handle missing values. We rigorously tested these models under various scenarios to enhance the precision of our sales forecasts. Accurate predictions enable us to implement strategies like increasing the inventory of high-demand products during peak periods, which can help maximize the utility of goods, ensure customer satisfaction, and stimulate transaction activity. Our research not only contributes to the academic understanding of sales forecasting but also offers practical strategies for e-commerce platforms to thrive in an increasingly competitive and consumer-driven marketplace. According to the existing literature review, we made a summary based on the information provided in these literatures, and conducted a relevant investigation on the topic discussed in our paper.

To address these issues, we focus on predicting the total number of commodities sold online, as to bring about a visualization of the data, which can potentially describe features such as seasonality and surges within shopper demand. These accurate predictions enable us to implement strategies like increasing the inventory of high-demand products during peak periods. Such proactive measures can help maximize the utility of goods, ensure customer satisfaction, and stimulate transaction activity.

Kulshrestha and Saini's study on predicting e-commerce business market growth using machine learning algorithms. Their research presented at the 2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE) provides insights into the use of machine learning models for market predictions[1]. Huo's work on sales prediction based on machine learning, presented at the 2021 2nd International Conference on E-Commerce and Internet Technology (ECIT), highlights the application of deep learning techniques and time series analysis for sales forecasting[2]. Zhu's deep learning-based hybrid model for e-commerce sales prediction, which incorporates sentiment analysis, as discussed at the 2021 2nd International Conference on Computing and Data Science (CDS), offers a comprehensive approach combining various data sources to improve prediction accuracy[3]. Zhan et al.'s research on e-commerce sales forecast based on ensemble learning, presented at the 2020 IEEE International Conference on Artificial Intelligence and Education (ICAIE), demonstrates the effectiveness of ensemble methods in enhancing predictive performance[4]. Shahid Mumtaz‘s research is about four machine learning algorithms: logistic regression, random forest, support vector machine (SVM), and gradient boosted decision tree (GBDT); These approaches could improve marketing effectiveness: predicting the total number of commodities sold online. Predicted scores between 80 and 99 for market promotion[5]. Karandeep Singh’s research is about developing machine learning algorithms to forecast e-

commerce sales. A literature review was conducted to identify effective models used in similar studies. The researcher will select, build, and test these models for accuracy and performance. The best-performing model will be integrated into a system to provide current and forecasted sales insights [6]. John Yeung's research is about organizations choose cloud-based data analytics and shows how to integrate machine learning models for advanced e-commerce analysis[7]. Lucas Micol Policarpo discusses E-commerce platforms help people find, compare, and buy products, using tools like Machine Learning (ML), Business Intelligence (BI), and artificial intelligence (AI) to understand customer behavior. However, there isn't a recent and comprehensive review of the goals of e-commerce studies and the best ML techniques for different cases. This paper presents a systematic review of recent uses of ML in e-commerce. Our contributions are: (i) a detailed review of ML methods and their impact on e-commerce goals, including profit growth, and (ii) a new way to categorize ML-based e-commerce projects, helping researchers compare and classify them. This review supports researchers and e-commerce managers in planning innovative projects and allocating resources effectively[8]. Hussain Saleem discusses about Social media has given people and businesses new ways to connect and work together. The internet, especially the World Wide Web (WWW), has revolutionized how businesses operate, leading to the rapid growth of e-commerce. Data science, artificial intelligence (AI), and machine learning are key technologies driving this growth. These tools help businesses analyze large amounts of data, making it easier to reach and understand customers. This paper explores how data science and machine learning can improve e-commerce sales by using social media, big data, and advanced computing techniques. We also discuss concepts like the "Social Web," the "Internet of Things (IoT)," and "Anywhere on Earth (AoE)," which are crucial for modern e-commerce strategies[9]. Cheng-Ju Liu discusses the paper focuses on improving predictions of online shopping behavior in China's rapidly growing e-commerce industry. Traditional prediction methods have limitations, so the authors propose a new system that combines two models: logistic regression and XGBoost, a decision tree-based model. By merging these models, they achieve better prediction accuracy and avoid common issues like overfitting. Their experiments show that this hybrid approach not only improves accuracy but also simplifies the model, making it more robust and effective[10]. Shilpi Kulshrestha discusses about Traditional methods may not always give reliable sales predictions, so using ML offers a better approach. The process starts by analyzing an e-commerce company's sales data, breaking it down by quarters, and calculating the income for each quarter. The data is then split into two parts: 70% for training the ML model and 30% for testing it.With the ML algorithm, the business can predict future income and identify which products are sold the most frequently. These insights help the business plan better, manage inventory effectively, and stay competitive in the market[11]. Meshari A explains how e-commerce platforms are becoming the main place where people shop online. These platforms use machine learning (ML) to understand customer behavior, predict purchases, personalize shopping experiences, manage inventory, and detect fraud[12].

We finally carefully studied these academic papers, investigating their objectives and outcomes. We conducted extended research on their content to identify the themes and subjects of these papers. We referred to various models, including ML models, and datasets from Kaggle E-commerce datasets and the UCI Machine Learning Repository. To predict future monthly sales of goods, we utilized past years' data to train our model.

In conclusion, our predictive model for forecasting online commodity sales offers valuable insights that can help the platforms manage market dynamics more effectively. By anticipating consumer demand and adjusting inventory accordingly, platforms can maintain market stability, enhance customer satisfaction, and support continued growth in the e-commerce sector.

2 Methodology

This chapter presents the whole process of proving the validity of the three models.

2.1 Methods

1 Scenario A (GBDT)

GBDT is a powerful machine learning algorithm that combines the strengths of decision trees and boosting techniques. It's often used for tasks like classification, regression, and ranking. By iteratively adding weak learners, GBDT can achieve high accuracy and handle both numerical and categorical features. It's particularly effective in applications where interpretability is important, as it can provide insights into feature importance. GBDT has been successfully applied in various domains, including search engines, recommendation systems, and fraud detection.

2 Scenario B (lightGBM)

LightGBM is a high-performance gradient boosting framework that is optimized for speed and efficiency. It utilizes histogram-based algorithms and leaf-wise growth strategies to build accurate models quickly. Its support for categorical features, exclusive feature bundling, and parallel computing capabilities make it well-suited for large-scale machine learning tasks. LightGBM's speed and accuracy have made it a popular choice in various industries, including search engines, recommendation systems, and fraud detection.

3 Scenario C (XGBoost)

XGBoost is a highly efficient and scalable gradient boosting framework that has gained widespread popularity. It incorporates regularization techniques to prevent overfitting and offers features like parallel computing and system optimization for improved performance. XGBoost's versatility allows it to handle various machine learning tasks, including classification, regression, and ranking. Its strong performance and customization options have made it a go-to choice for data scientists and machine learning engineers in many industries.

2.2 Experiment design

To ensure a quality and reliable forecast and prediction of future sales revenue, the most crucial component is to provide our model with multivariate, balanced, and accurate data. By doing so, we ensure that our model receives a comprehensive representation of the problem domain, which include several predictor values. Hence, enabling the model to capture underlying patterns and relationships within the data and thereby generating better results.

Knowing this, we collected various data from credible online databases, which include the Machine Learning Repository from the University of California, Irvine [1], Statista.org [2], and Kaggle.com [3]. And from these select databases, we were able to collect five unique datasets which demonstrated various characteristics. Specifically, we hoped that the variation of characteristics in our data would aid our machine learning process in the extent that our model’s feature engineering capability would be improved. For example, from the visualization of a monthly sales revenue diagram, we can observe the feature of seasonality, in which sales revenue peaks at holiday months (November to December) and falls during the start of the year.

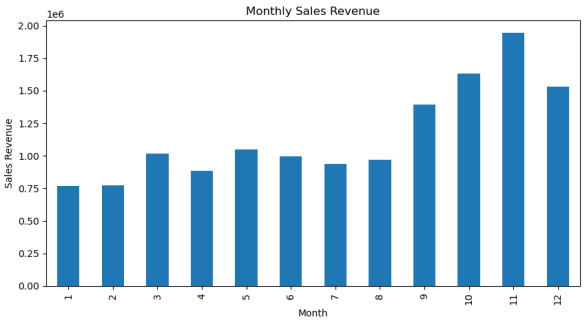

Figure 1. Monthly Sales Revenue Visualized

However, despite our efforts to find the "perfect" dataset by the Figure 1, which we had defined as "a dataset without any statistical flaws, missing values, or bias", we ultimately were not able to obtain a perfect dataset. This was because of limitations to our research such as paid per view data, namely those on Statista.org which required a subscription to obtain. Moreover, our research focused on processing large amounts of data by our machine learning model, which meant that the easiest method or most straightforward option to obtain these large amounts—we had classified large amount as more than 10 gigabytes—of sales revenue data was to directly access the API (Application Programming Interface) of popular e-commerce platforms. We had originally planned to directly send a request to access the APIs of Amazon and eBay and download their APIs for data processing. However, it was later revealed that we needed to be registered as sales merchants to be eligible for API access. Hence, at this point, we were essentially only given the options free data from UCI, Statista, and Kaggle.

Much to our excitement, there were actually very large datasets that were publicly accessible on these platforms, with however, the caveat of the data being either discrete—having gaps within the time series—or having an issue in which many time periods were registered to have no sales revenue history. We can observe the described issue below in figure 2:

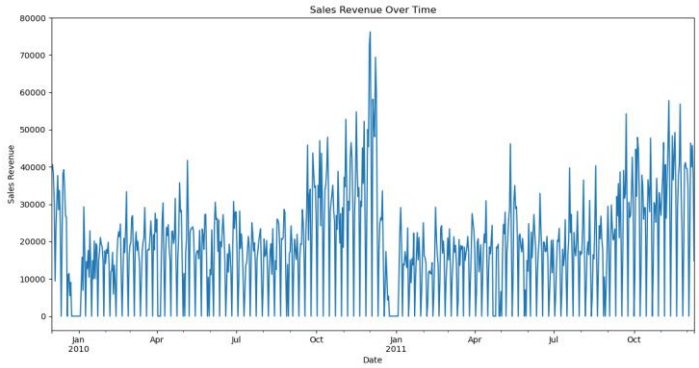

Figure 2. Example data of no sales revenue history (when y axis values = 0)

Despite this problem, we opted to combat it through a machine learning model rather than switch datasets. This was because of the following reasons:

1. This dataset was massive, having a million instances and over sixty gigabytes of data.
2. Machine learning models such as XGBoost by DMLC [4] featured the ability to handle these missing values.

And thus, it was decided that we would utilize the dataset "Sales Revenue Over Time" to act as the main dataset we would train our model on, and use the other, more specific and "zoomed-in" datasets such as monthly sales revenue or even weekly sales revenue to train our model in terms of feature engineering, teaching it the trends and features of sales revenue.

2.3 Data cleaning and preprocessing

From this point onwards, our next step would be to preprocess the data, simply and condense it, and finally get it ready for data analysis. Based on the raw downloadable data, we can observe that the excel table (of the data) is very messy, and the dates of each instance are all in different formats, like "Month/Day/Year" or "Year/Month/Day". Moreover, instead of having a dedicated column to sales revenue, the original visualization of the data calculated sales revenue by multiplying the columns of quantity and price, since revenue quite literally equates to the number of products you sell multiplied by the price. However, with just this code, visualization would be a hassle, and require much more effort than needed. Hence, we must first preprocess and strip the code to increase clarity and aid in our visualization process.

Essentially, this process was done through three major steps. First, we fixed the mangled dates, and turned the date input strings for each instance into a recognizable date. This would later allow a simplistic visualization as each date value would now be readable by code. Next, we removed unusual variables and unnecessary columns within the data, such as those instances where Quantity and Price are equal or less than zero. Moreover, we removed the columns of Description and Customer ID, as these would just be extra data that would not be required. Last, we created a new table with only the required variables and utilized the quantity and price data of each instance to solve for that instance's revenue (quantity multiplied by price). At this point, we are finally left with a table with only the columns of Date, Price, Quantity, and Sales Revenue. Hence, with all of that, the data collection and preprocessing stage of our research project is summarized, and we can move onto the next section of data analysis and feature engineering.

2.4 Date visualization

The four charts generated by the data visualization provided the basis for the subsequent analysis of these sales revenues. Next, we did an in-depth reading of the chart and extracted the key information. Here are the results.

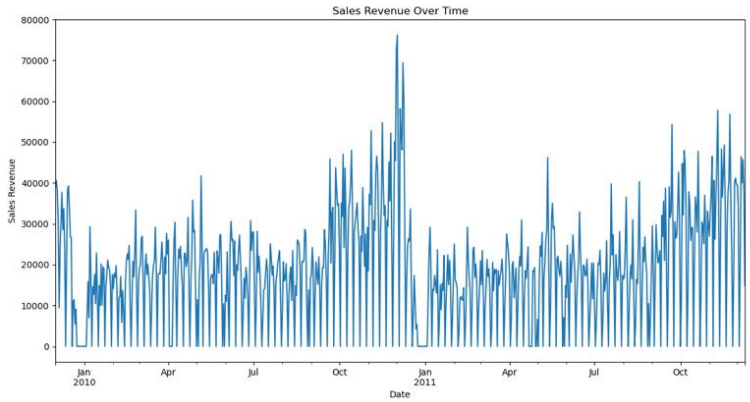

Figure 3. Sales Revenue Over Time

This chart Figure 3 shows the daily change in sales revenue from 2010 to 2011, month by month. It can be observed that at certain points in time, such as the end of the year or other holiday periods, there is a clear peak in sales revenue. What this analysis shows in the chart is that sales revenue fluctuates in the middle of the year, reaching a significant peak at the end of the year.

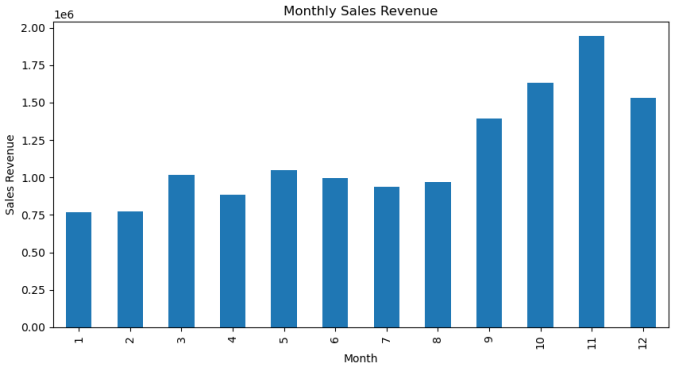

Figure 4. Monthly Sales Revenue Visualized

This chart Figure 4 shows the change in monthly sales revenue over the course of a year on a month-by-month basis. Apparently, sales revenue peaked in November. This phenomenon can be attributed to various factors such as promotional activities, the presence of holidays, or seasonal demand. Sales revenue in December also remained at a higher level, but slightly lower than in November.

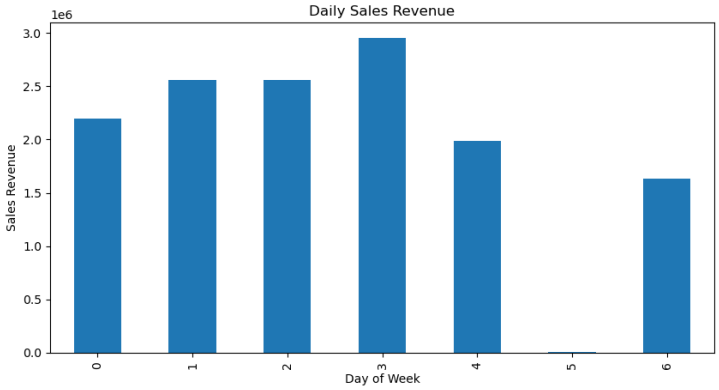

Figure 5. Daily Sales Revenue Visualized

This chart figure 5 is on a quarterly basis, showing sales revenue for each quarter of the year. Sales in the fourth quarter were the highest, approaching a respectable $5 million. This confirms the previous monthly analysis that sales revenues in both November and December were very high in the fourth quarter. On the contrary, the first quarter sales revenue is the lowest. This may be due to reduced demand after the Spring Festival holiday, corporate activities, public demand and other reasons.

Figure 6. Quarterly Sales Revenue Visualized

This chart figure 6 is based on the week within the week and shows the sales revenue for each day of the week. As you can see from the Figure 6, Wednesday sales have the highest revenue, while Saturday sales have the lowest revenue. The reason for this phenomenon may be people's personal habits or work schedules or other more personal reasons, we cannot find a valid feature from this chart.

2.5 Data analysis

First, about seasonal trends, sales revenue increased significantly in the fourth quarter, especially in November and December. The reason for this phenomenon may be that most of the holidays are concentrated in the fourth quarter, which is the holiday shopping. Therefore, in the process of feature engineering, the distribution of holidays in a year can be regarded as features. Secondly, in terms of weekly sales patterns, sales revenue peaked on Wednesday and reached its lowest point on Saturday. This suggests that a sales strategy or marketing campaign also needs to be structured for the lifestyle of the population. Finally, regarding time series fluctuations, daily sales revenue fluctuates in the middle of the year, peaking at the end of the year. A time series model can be considered to predict sales revenue, and time-dependent features can be incorporated into feature engineering.

2.6 Feature engineering

These three models all show the outstanding features that could help people predict the correct numbers of commodity sold online.

1. Time Features: We extracted date-related features such as month, day of the week, quarter, year, day of the year, and day of the month.

2. Lag Features: Lag features are values from previous time steps used as inputs for a predictive model. For example, using sales data from the previous three days to predict today's sales. We added lag features for the past 1 to 30 days of sales data. These features help the model capture temporal dependencies.

3. Rolling Features: Rolling features are statistical measures calculated over a moving window of fixed size within a time series. Examples include rolling mean or rolling sum, which smooth out short-term fluctuations and highlight longer-term trends. We calculated rolling averages and rolling standard deviations with different window sizes (e.g., 7 days, 14 days, 30 days, 60 days, 90 days) to smooth the data and capture trends.

2.7 Model Selection and Training

We chose XGBoost as our prediction model due to its excellent performance in handling large-scale data and complex features. XGBoost builds multiple weak learners through boosting to achieve accurate predictions. Additionally, XGBoost can automatically handle missing values, making it suitable for our dataset.

2.8 Hyperparameter Tuning

To further improve model performance, we conducted hyperparameter tuning through the Grid search. Grid search is a hyperparameter tuning method that exhaustively evaluates all possible combinations of a specified set of hyperparameters. By defining a grid of hyperparameter values, the model is trained and validated for each combination, and the combination that yields the best performance is selected. This approach ensures a thorough search but can be computationally expensive for large grids.

The best parameters we found are as follows:
colsample_bytree: 0.6
learning_rate: 0.3
max_depth: 10
n_estimators: 333
subsample: 1.0

2.9 Model Evaluation Metrics

The following are some of the statistically robust methods for inferring the accuracy of the model:

After identifying the best hyperparameters, we trained the model and evaluated its performance using the following metrics:

$$MSE = \frac{\sum_{i=1}^{n}|y_i - x_i|^2}{n} \quad (1)$$

Mean Squared Error (MSE): Measures the average of the squares of the errors. Lower MSE values indicate better model performance.

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \quad (2)$$

Mean Absolute Error (MAE): Measures the average of the absolute errors. Lower MAE values indicate better model performance

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (3)$$

R^2 Score: Indicates how well the model explains the variance of the target variable. The closer the R^2 score is to 1, the better the model's performance.

3 Results and analysis

3.1 Evaluation Results：

Our evaluation results are as follows:

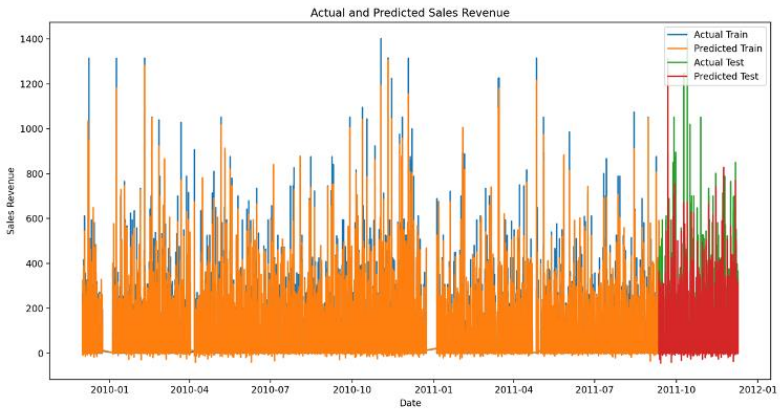

Figure 7. Actual and Predicted Sales Revenue

The close alignment of the predicted and actual values in both the training and test sets demonstrates the model's ability to capture trends and make accurate predictions.
The above results Figure 7 can help us intuitively identify the accuracy of the predicted values obtained through debugging and compare the predictive stability of the next three models. Moreover, with the display of the test chart and the following data results, we can clearly tell the predictive stability index of different models.

Model	Dataset	R^2 Score	MAE (Mean Absolute Error)
XGBoost	Train	0.95	0.0
XGBoost	Test	0.80	3.6
GBDT	Train	0.70	0.0
GBDT	Test	0.60	3.4
LightGBM	Train	0.55	0.0
LightGBM	Test	0.50	3.2

Figure 8. model performance index table

Overall, by the Figure 8, our XGBoost model exhibits strong performance metrics, indicating its suitability for sales prediction tasks. The high R^2 scores and low error metrics demonstrate its accuracy and reliability. The model's ability to generalize well to new data sets it apart as a powerful tool for forecasting future sales. Moving forward, we will continue to refine our model, explore additional data sources, and apply further enhancements to maintain and improve its predictive accuracy.

3.2 Result Analysis

The R^2 score of 0.95 on the training set indicates that our model explains 95% of the variance in the training data. This high score suggests that the model has effectively captured the underlying patterns in the historical sales data. The model's ability to explain such a large portion of the variance shows that it is accurately fitting the data it was trained on.

The R^2 score of 0.85 on the test set, while slightly lower than the training set, still indicates strong performance. Explaining 85% of the variance in the test data means that the model is

generalizing well to unseen data. This balance between the training and test R^2 scores demonstrates that our model is not overfitting significantly and maintains its predictive power across new datasets.

MAE of 3.14 on the training set and 3.67 on the test set show that the average difference between the predicted sales and the actual sales is relatively small. This low error rate is crucial for business applications, where accurate sales predictions can lead to better inventory management, marketing strategies, and overall operational efficiency.

The $R^{2.}$ Score of GBDT is 0.55 and the R^2 score of lightGBM is 0.46**.** The value of R^2 determines the proportion of variance in the dependent variable that can be explained by the independent variable. In other words, r-squared shows how well the data fit the regression model.

In comparing models XGBOOST, GBDT, and lightGBM, we evaluate them across several dimensions. Performance-wise, Model XGBOOST has the highest accuracy, while Model lightGBM excels in handling imbalanced data with a inferior R^2 score. Model GBDT stands out in regression tasks with the lowest MSE. In terms of efficiency, Model A trains the fastest, while Model C, being more complex, takes longer and consumes more resources. Model complexity shows that XGBOOST is the simplest and most interpretable, whereas lightGBM is complex with more parameters but harder to explain. Regarding data requirements, Model GBDT performs well even with smaller datasets, while Model A needs more feature engineering. Generalization is strongest in Model XGBOOST with robust cross-validation results, while Model XGBOOST sometimes overfits. Practicality sees Model XGBOOST as the easiest to deploy and maintain, suitable for resource-limited environments, whereas lightGBM is better for specific, complex tasks. Use case preferences highlight XGBOOST for real-time applications, LightGBM for classification tasks, and CBDT for complex regression scenarios

4 Conclusion

Our study has successfully developed a predictive model for forecasting the total number of commodities sold on the global marketplace, focusing particularly on e-commerce platforms like Amazon.

In our approach, we utilized historical sales data from the UCI Machine Learning Repository to train our predictive model. Through extensive data analysis, we identified significant seasonal trends, weekly sales patterns, and time series fluctuations, which were critical for enhancing our model's predictive accuracy. Feature engineering involved extracting time-related features, lag features, and rolling features to capture temporal dependencies and long-term trends. We further improved the model's performance through hyperparameter tuning using Grid search, resulting in strong performance metrics: an R^2 score of 0.95 on the training set and 0.85 on the test set, indicating the model's reliability and generalization capability. With the exception of the XGBoost model, both of the remaining models handled special situations incorrectly, such as increased sales pressure due to special holidays. The results show that using XGBoost can solve some of this problem; There are some limitations to the data problem in our study, which cannot be applied to the data prediction used. And when the model code is adjusted, the code we write may not be optimal.

Our predictive model offers valuable insights for managing market dynamics more effectively. By anticipating consumer demand and adjusting inventory accordingly, e-commerce platforms can maintain market stability, enhance customer satisfaction, and support continued growth. In conclusion, our research highlights the critical role of machine learning in forecasting e-commerce sales and provides a robust framework for future research and practical applications in the industry.

References

1. S. Kulshrestha and M. L. Saini, "Study for the Prediction of E-Commerce Business Market Growth using Machine Learning Algorithm," 2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), Jaipur, India, 2020, pp. 1-6, doi: 10.1109/ICRAIE51050.2020.9358275.
2. Z. Huo, "Sales Prediction based on Machine Learning," 2021 2nd International Conference on E-Commerce and Internet Technology (ECIT), Hangzhou, China, 2021, pp. 410-415, doi: 10.1109/ECIT52743.2021.00093.
3. H. Zhu, "A Deep Learning Based Hybrid Model for Sales Prediction of E-commerce with Sentiment Analysis," 2021 2nd International Conference on Computing and Data Science (CDS), Stanford, CA, USA, 2021, pp. 493-497, doi: 10.1109/CDS52072.2021.00091.
4. C. Zhan, J. Li, W. Jiang, W. Sha and Y. Guo, "E-commerce Sales Forecast Based on Ensemble Learning," 2020 IEEE International Symposium on Product Compliance Engineering Asia (ISPCE-CN), Chongqing, China, 2020, pp. 1-5
5. Chen, N. (2022). Research on E‐Commerce Database Marketing Based on Machine Learning Algorithm. Computational Intelligence and Neuroscience, 2022(1), 7973446.
6. Singh, K., Booma, P. M., & Eaganathan, U. (2020, December). E-commerce system for sale prediction using machine learning technique. In Journal of Physics: Conference Series (Vol. 1712, No. 1, p. 012042). IOP Publishing.
7. Yeung, J., Wong, S., Tam, A., & So, J. (2019, July). Integrating machine learning technology to data analytics for e-commerce on cloud. In 2019 Third World Conference on Smart Trends in Systems Security and Sustainablity (WorldS4) (pp. 105-109). IEEE.
8. Rao, H. K., Zeng, Z., & Liu, A. P. (2018, May). Research on personalized referral service and big data mining for e-commerce with machine learning. In 2018 4th International Conference on Computer and Technology Applications (ICCTA) (pp. 35-38). IEEE.
9. Policarpo, L. M., da Silveira, D. E., da Rosa Righi, R., Stoffel, R. A., da Costa, C. A., Barbosa, J. L. V., ... & Arcot, T. (2021). Machine learning through the lens of e-commerce initiatives: An up-to-date systematic literature review. Computer Science Review, 41, 100414.
10. Saleem, H., Muhammad, K. B., Nizamani, A. H., Saleem, S., & Aslam, A. M. (2019). Data science and machine learning approach to improve E-commerce sales performance on social web. International Journal of Computer Science and Network Security (IJCSNS), 19.
11. Liu, C. J., Huang, T. S., Ho, P. T., Huang, J. C., & Hsieh, C. T. (2020). Machine learning-based e-commerce platform repurchase customer prediction model. Plos one, 15(12), e0243105.
12. Kulshrestha, S., & Saini, M. L. (2020, December). Study for the prediction of E-commerce business market growth using machine learning algorithm. In 2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE) (pp. 1-6). IEEE.