ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 10| October 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

PREDICTING PRODUCT DEMAND USING MACHINE LEARNING IN SUPPLY CHAINS

Kantha Lakshminarasimhan
Department of Data Science, Dr. D. Y. Patil College, Pimpri, Pune, India

ABSTRACT

Accurate demand forecasting is a critical challenge in modern retail supply chains, as it directly impacts inventory planning, cost optimization, and customer satisfaction. This study applies machine learning techniques to predict weekly sales using historical retail data. After preprocessing and exploratory data analysis, multiple models were evaluated, including Linear Regression, Random Forest, XGBoost, and LightGBM. Results showed that Linear Regression performed poorly with an RMSE of approximately 16,019, while tree-based models significantly improved accuracy. The tuned XGBoost model achieved the best performance with an RMSE of 4,552, representing a 72% improvement over the baseline. Feature importance analysis revealed that store ID, year, unemployment rate, and CPI were the most influential predictors of sales. Residual analysis confirmed stable predictions, and actual vs predicted plots demonstrated close alignment between forecasts and true values. While the model achieved strong results, limitations included restricted dataset features and computational requirements. Overall, this research demonstrates that machine learning, particularly XGBoost, provides an effective and scalable approach for demand forecasting in supply chains.

KEYWORDS: *Demand Forecasting, Supply Chain, Machine Learning, XGBoost, Prediction*

INTRODUCTION

The supply chain is the backbone of many industries, and effective demand forecasting is vital for maintaining inventory levels, reducing operational costs, and ensuring customer satisfaction. However, traditional forecasting techniques, such as moving averages and exponential smoothing, are often limited by their inability to model nonlinear and high-dimensional data relationships.

With the exponential growth in data and computational capabilities, machine learning has emerged as a transformative tool in supply chain analytics[1]. ML models can learn from historical data and uncover hidden patterns that influence demand, such as seasonal trends, promotions, and external events.This research investigates the application of ML models for improving product demand forecasting and addresses key challenges like data preprocessing, feature selection, and model evaluation.

RELATED WORK

Earlier research in demand forecasting was largely based on statistical models such as ARIMA and exponential smoothing[2]. These methods worked well for linear and seasonal patterns but struggled with non-linear dependencies and external factors. Later studies incorporated machine learning models such as Decision Trees, Random Forests, and Gradient Boosting[3]. More recently, XGBoost and LightGBM have become popular due to their scalability and accuracy[4]. Deep learning approaches such as LSTM networks have also been tested[5] but require large datasets and high computational resources.

PROBLEM STATEMENT

Supply chains frequently suffer from inaccurate demand forecasts, leading to excess inventory, stockouts, and lost revenue. Traditional models lack the adaptability and precision required for dynamic and complex environments.

OBJECTIVES

1. To analyze historical sales and supply chain data to identify relevant trends and demand patterns.
2. To build and compare multiple ML models (e.g., Random Forest, XGBoost, LSTM) for product demand forecasting.
3. To evaluate the effectiveness of each model based on predictive accuracy and real-world applicability.

1.METHODOLOGY

1.1 Data Collection

The dataset for this study was sourced from real-world and open platforms such as Kaggle's *Retail Demand Forecasting* dataset and the UCI Machine Learning Repository. It includes historical sales data along with external factors such as promotions, seasonal indicators, and calendar events, which provide context for understanding demand fluctuations.

1.2 Data Preprocessing

Preprocessing steps were applied to ensure data quality and consistency. Missing values were handled appropriately, outliers were detected and addressed, and numerical features were normalized to a standard scale. Categorical variables, such as holiday indicators, were encoded using one-hot encoding to make them suitable for machine learning models.

1.3 Feature Engineering

Feature engineering was performed to enhance the predictive power of the models. Time-based features such as day, week, and month were extracted to capture seasonal patterns. In addition, lag features and rolling averages of past sales were created to incorporate temporal dependencies into the forecasting process.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 10| October 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

1.4 Modelling Techniques

Several machine learning models were employed for demand forecasting. As baselines, Linear Regression and Decision Trees were implemented. Advanced models included ensemble methods such as Random Forest and Gradient Boosting (XGBoost), which are capable of capturing non-linear relationships and complex interactions. Furthermore, Long Short-Term Memory (LSTM) networks were considered to exploit the sequential nature of time-series data.

1.5 Evaluation Metrics

The models were evaluated using standard regression metrics. Root Mean Square Error (RMSE) was used to measure the average magnitude of prediction errors, while Mean Absolute Error (MAE) provided insight into the average absolute difference between predictions and actual values. Additionally, Mean Absolute Percentage Error (MAPE) was used to express prediction accuracy as a percentage, allowing for easier interpretability across different scales.

2. RESULTS

Several visualizations were used to better understand the dataset:

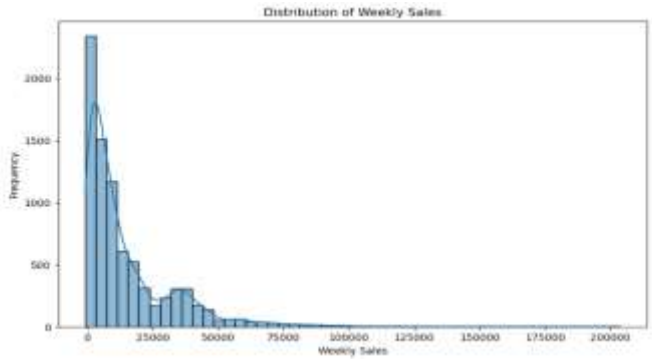

Figure 1: The distribution of weekly sales is right-skewed, with most sales clustered at lower values and a few extreme peaks.

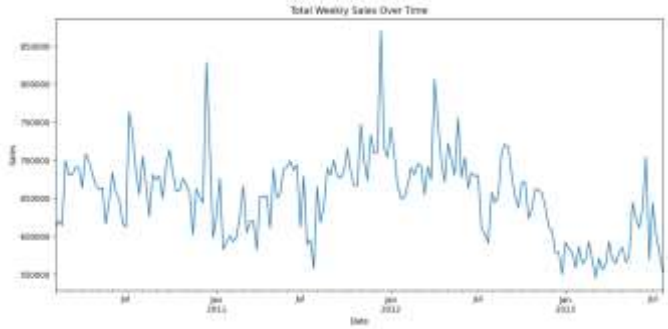

Figure 2:Illustrates the weekly sales trend over time, showing seasonal fluctuations and noticeable spikes during holiday periods.

.

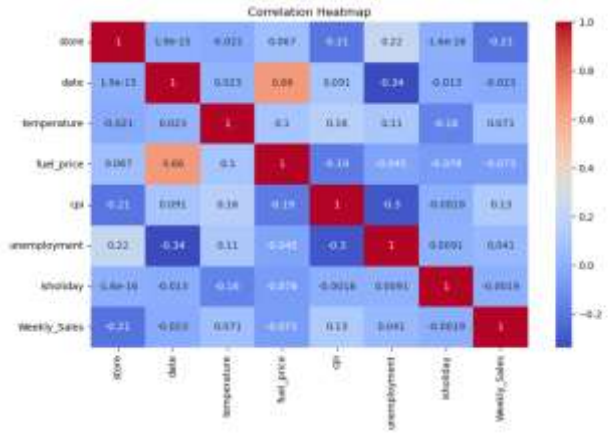

Figure 3:The relationships between features are summarized, where CPI and unemployment show moderate correlation with sales, while temperature and fuel price are weakly correlated

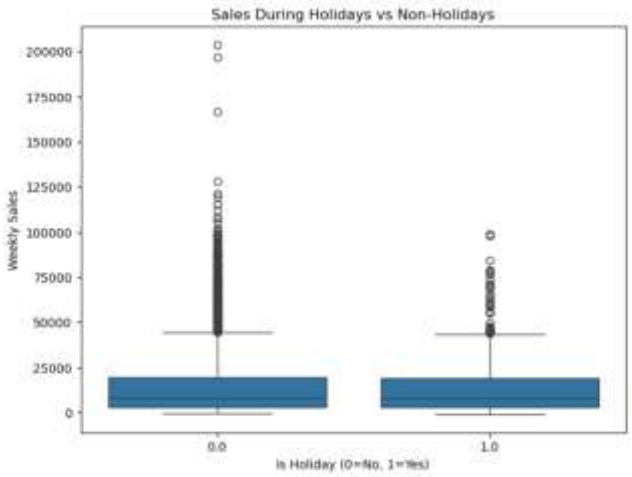

Figure 4: Average weekly sales are significantly higher during holiday weeks compared to non-holiday weeks.

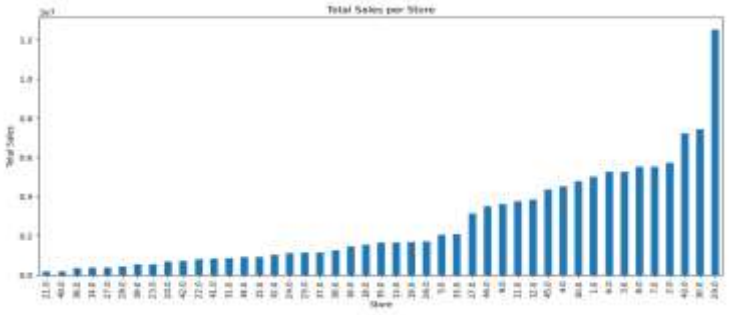

Figure 5:Compares the average *weekly sales across different stores, highlighting the variation in demand patterns by location.*

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 10| October 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

Table 1: RMSE Comparison Across Models (Linear Regression, Random Forest, XGBoost, LightGBM).

Model	RMSE
Linear Regression	16018.81
Random Forest	5074.39
XGBoost	4996.63
LightGBM	5157.56
Tuned XGBoost	**4551.94**

- After hyperparameter tuning, XGBoost achieved the best performance with an RMSE of 4551.94, outperforming other models.

The performance of different models is summarized in **Table 1**, where tuned XGBoost achieved the lowest RMSE, outperforming all other approache

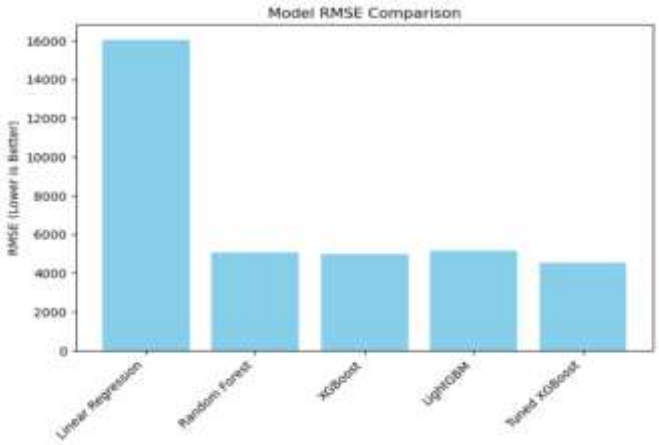

Figure 6: Ensemble models such as Random Forest, XGBoost, and LightGBM significantly reduced RMSE compared to Linear Regression, with tuned XGBoost achieving the best performance.

Other performance metrics were calculated for the Tuned XGBoost model:

- MAE (Mean Absolute Error)**:** measures average absolute deviations.
- R^2 Score: high values indicated strong explanatory power.
- MAP
- E (Mean Absolute Percentage Error)**:** predictions were within a small percentage error.

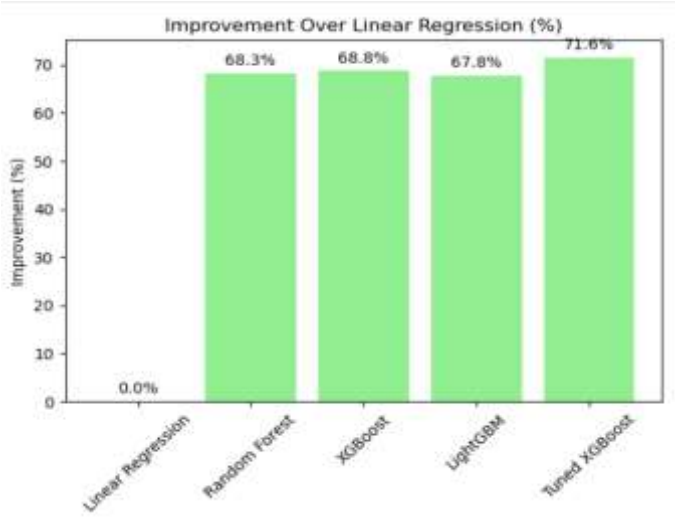

Figure 6:Illustrates the percentage improvement of advanced machine learning models over Linear Regression, where tuned XGBoost achieved the highest reduction in error, followed by Random Forest and LightGBM.

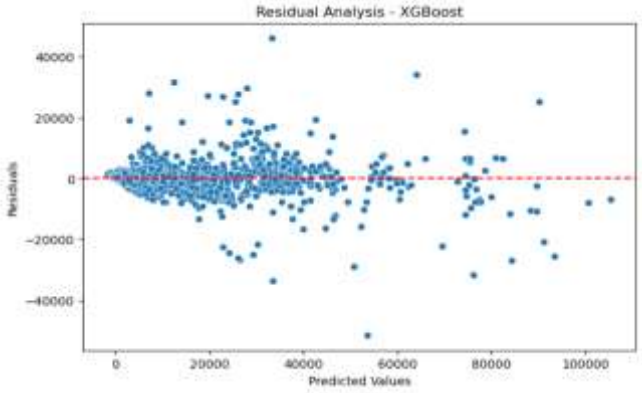

Figure 7:The residuals versus predicted plot shows no clear pattern, confirming that errors are randomly distributed and the model generalizes well.

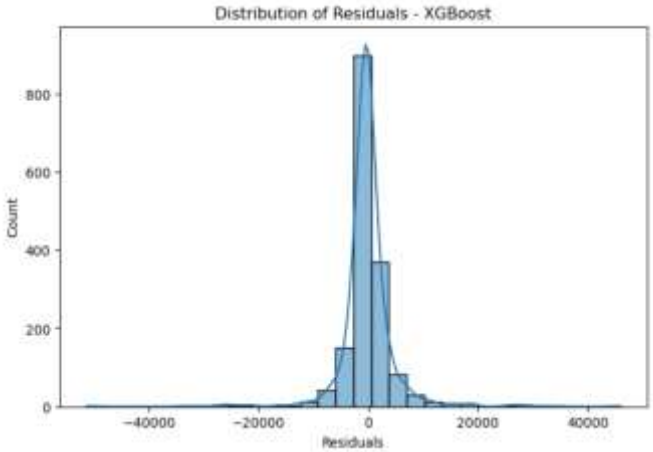

Figure 8:The residuals of the tuned XGBoost model are approximately normally distributed and centered around zero, indicating that the model errors are unbiased.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 10| October 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

Figure 9:The predictive accuracy of the tuned XGBoost model is demonstrated, where actual sales values closely align with predicted values.

- **Feature Importance**

The tuned XGBoost model identified the following most influential features:

1. Store ID (30.8%)
2. Year (21.9%)
3. Unemployment Rate (17.9%)
4. CPI (15.3%)

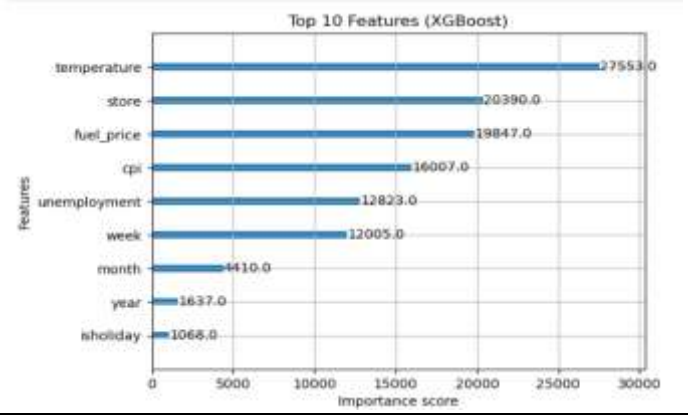

Figure 10:Feature importance rankings are displayed, where store, year, unemployment, and CPI emerge as the most influential predictors of sale.

DISCUSSION

Tree-based models significantly outperformed linear models due to their ability to capture non-linear relationships. XGBoost's performance was enhanced through hyperparameter tuning. Feature importance insights can guide inventory and store-level decisions. Limitations include dataset scope and computational cost.

CONCLUSION

Machine learning techniques, especially XGBoost, offer scalable and accurate solutions for demand forecasting. This approach supports better inventory planning and cost reduction. Future work should incorporate richer datasets and explore deep learning models for long-term forecasting.

ACKNOWLEDGMENT

The author thanks Dr. D.Y. Patil College for guidance and support.

CONFLICT OF INTEREST

The author declares no conflict of interest.

REFERENCES

1. *Chopra S, Meindl P. Supply Chain Management: Strategy, Planning, and Operation. Pearson Education; 2016.*
2. *Carbonneau R, Laframboise K, Vahidov R. Application of machine learning techniques for supply chain demand forecasting. Eur J Oper Res. 2008;184(3):1140–1154.*
3. *Makridakis S, Spiliotis E, Assimakopoulos V. Statistical and machine learning forecasting methods: Concerns and ways forward. PLOS ONE. 2018;13(3): e0194889.*
4. *Zhang G, Patuwo BE, Hu MY. Forecasting with artificial neural networks: The state of the art. Int J Forecasting. 1998;14(1):35–62.*
5. *Aburto L, Weber R. Improved supply chain forecasting using artificial neural networks and ARIMA. Expert Syst Appl. 2007;29(1):135–144.*