ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr2260

Role of Artificial Intelligence (AI)-Driven Demand Forecasting: A Machine Learning Approach for Supply Chain Resilience

Prateek Bansal[1]

[1] MBA Student,University of Illinois, Urbana Champaign

Publication Date: 2025/05/17

Abstract: Reducing financial risks, improving inventory management, and strengthening supply chain resilience all depend on accurate demand forecasts. Traditional forecasting methods often struggle with unpredictable market fluctuations, seasonal variations, and external disruptions, leading to inefficiencies such as stockouts and overstocking. This study leverages artificial intelligence (AI) and machine learning techniques to improve sales prediction accuracy using real-world Walmart sales data. This study utilizes ML techniques to predict sales accurately, comparing XGBoost, LightGBM, Random Forest, and K-Nearest Neighbors (KNN). A methodology involves data preprocessing, including data cleaning, one-hot encoding, and normalization, followed by feature selection and dataset splitting. XGBoost and LightGBM models outperform traditional methods, achieving high R^2 values of 0.9752 and 0.9732, respectively, with low MSE, RMSE, and MAE, indicating strong predictive capabilities. Comparative analysis reveals that Random Forest (R^2 = 0.9569) and KNN (R^2 = 0.9381) exhibit lower accuracy. The actual vs. predicted sales plots for XGBoost and LightGBM demonstrate close alignment, while residual plots confirm minimal bias. Overall, the findings highlight the superiority of gradient boosting techniques in demand forecasting, offering valuable insights for effective sales prediction and inventory planning in the retail sector.

Keywords: *Demand Forecasting, Machine Learning, Supply Chain Resilience, Walmart Sales Data, AI-Driven Decision Making.*

How to Cite: Prateek Bansal(2025). Role of Artificial Intelligence (AI)-Driven Demand Forecasting: A Machine Learning Approach for Supply Chain Resilience. *International Journal of Innovative Science and Research Technology*, 10(4), 4056-4067. https://doi.org/10.38124/ijisrt/25apr2260

I. INTRODUCTION

Through linked procedures, SCM entails the smooth transfer of information, products, and services from suppliers to final consumers [1]. The classical supply chain management models rely on static predetermined values for capacity and demand as well as cost. Actual supply chain operations exhibit high dynamic conditions because of unpredictable customer demand along with supply chain interruptions and transportation barriers and organizational vulnerabilities [2][3][4]. Demand unpredictability stands as the most important supply chain performance issue, which affects essential operational areas, including logistics, together with inventory planning and production scheduling. The situation requires demand forecasting to cut down uncertainty and strengthen decision-making performance. Capacity, demand, and cost are often considered to be known features in SCM challenges [5][6][7][8]. However, in reality, there are a lot of unknowns due to factors like supply transportation, customer demand, organizational risks, and variations in lead times, therefore this doesn't work [9]. Demand uncertainties create the largest negative effects on SC performance while triggering widespread impacts across transportation systems and inventory management and production scheduling. Therefore, a vital tactic for fixing supply chain problems is demand forecasting [2][10].

E-commerce requires exact demand forecasting because changing consumer patterns result from seasonal trends and promotional campaigns and market movements. The complexity of present markets remains hidden to traditional forecasting techniques that utilize past sales data together with basic statistical models [11][12]. Therefore, it is convenient to apply more progressive trends for demand forecasting and SCM in businesses. AI and ML have revolutionized the method of demand forecasting through trends, patterns, and linkages between huge datasets which are not easily identifiable by normal models [13][14]. Companies may decide to improve the existing models with multiple elements such as sales history, market dynamics, time, and seasonal factors as well as promotion influences in a bid to improve an efficiency of their supply chains and boost forecasts [15]. SCM proactive tactics are well-suited to AI-driven demand forecasting, much as it is used in customer retention industries [16].

The advantages of ML on enhanced demand predictions are discussed in this study as a consequence of its application to SCM and AI in the context of demand forecasting. This study aims to review AI strategies, challenges, and possible

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr2260

future advancements in order to illuminate how businesses may use AI for robust and efficient SCM. This research mostly contributes to the following areas:

- Utilization of Walmart sales data, which provides real-world transactional records for demand forecasting.
- Enhancing data quality through data cleaning, one-hot encoding, and normalization, ensuring more reliable model performance.
- Feature selection and extraction to identify critical factors influencing sales trends, optimizing model efficiency.
- Implementation of XGBoost and LightGBM for accurate demand forecasting, leveraging their gradient boosting capabilities to improve prediction accuracy.
- Comparative analysis of ML models, including RF and KNN, evaluated using R^2, RMSE, MAE, and MSE to determine the most effective forecasting approach.
- Strengthening supply chain resilience by enabling predictive insights that reduce stock imbalances, enhance inventory optimization, and minimize financial risks.

➢ *Motivations of the Study*

The growing importance of precise and intelligent demand forecasting in contemporary SCM is driving this research. Traditional forecasting methods, such as statistical and heuristic-based approaches, often struggle with dynamic market conditions, seasonal variations, and unforeseen disruptions, leading to inefficiencies in inventory management, stockouts, and excess supply. Applying ML algorithms to historical sales data, identifying patterns, and producing precise demand projections, AI-driven demand forecasting enhances operational efficiency and resilience of the supply chain. To enhance an accuracy of demand projections, this study aims to use state-of-the-art ML approaches.

➢ *Novelty of Paper*

The novelty of this paper lies in the application of XGBoost and LightGBM for AI-driven demand forecasting, leveraging real-world Walmart sales data. The study enhances forecasting accuracy through comprehensive data preprocessing, including data cleaning, one-hot encoding, and normalization, along with feature selection to identify key demand-influencing factors. A comparative analysis against traditional models, such as RF and KNN, highlights the superior performance of gradient boosting techniques. Additionally, this research explores the potential for real-time demand forecasting, enabling proactive inventory management and strengthening supply chain resilience against disruptions.

➢ *Organization of the Paper*

The paper begins with Section II, which reviews existing research on traditional and machine learning-based forecasting models. Section III details data preprocessing, along with evaluation metrics. Section IV compares models with visualizations of performance metrics. The most important results, difficulties, and areas for further study are summarized in Section V.

II. LITERATURE REVIEW

The literature review explores AI-driven demand forecasting in supply chain resilience.

Areerakulkan et al. (2024) propose a method to improve demand forecasting for new mobile phones, leading to cost reduction and more efficient inventory management. Using Holt-Winters, ARIMA, ETS, and ANN, the most accurate method is selected, with ANN being 51.28% more accurate. The proposed solution plan reduces loss sales and inventory carrying costs by 27.71%.[17]

Lei et al. (2024) offer a transfer learning framework that can be used to create a bottom-level prediction model by merging data from several aggregation layers. By regularizing high-level sales data, the model outperforms JD.com's benchmark technique by 9 percent when it comes to predicting. They validate its generalizability using a Walmart retail data set in addition to other methods of pooling and prediction. With more accurate predictions, low-margin e-commerce businesses may minimize operating expenditures by 0.01-0.29 RMB per sold item, a considerable savings. [18]

Yani and Aamer (2023) investigate how pharmaceutical SCM makes use of ML methods for demand forecasting. The Konstanz Information Miner platform was used to examine three experimental designs utilizing a single-case explanatory technique. The results showed that simple tree and RF performed better than other algorithms, leading to a 41% increase in the accuracy of demand forecasting[19]

Shibu and Agarwal (2023) aim to deliver the importance of intervening in the resources and abilities of data analysis and visualization to accurately determine the expected sales values with the SCM process for improving the overall business. A dataset with record of five years of actual sales data is used. On this data four prediction models- Seasonal Naïve, Holt-Winters Triple Exponential Smoothing, Seasonal ARIMA and Linear Regression were trained and their results were analyzed. Each model focused on its characteristic way of determining outputs and was capable of overall delivering good results with an error rate not exceeding 27%.[20]

Ma et al. (2023) investigate demand forecasts for new energy vehicles employing a novel combination model called SARIMA- LSTM- BP. When compared to standard econometric and DL models, the model performs better in terms of accuracy and forecasting, as shown by its reduced RMSE, MSE, and MAE values. The creation of new energy vehicles depends on this.[21]

Taghiyeh et al. (2023) advise a hierarchical supply chain that minimizes transportation expenses, particularly for e-commerce sites, by improving parent-level forecasts with child-level predictions. It suggests a multi-stage machine learning strategy that improves prediction accuracy by 82% to 90% and helps planners in the supply chain make better use of multivariate data[22]

Table 1 summarizes research on AI-driven demand prediction and supply chain optimization, highlighting diverse methodologies. Key findings include improved forecasting accuracy, cost reduction, and operational efficiency, while challenges involve data quality, implementation costs, and real-world validation.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr2260

Table 1 Summary of Reviewed Works on AI-Driven Demand Forecasting in Supply Chain Resilience

Reference	Technology/Method Used	Data Characteristics	Key Findings	Advantages	Limitations & Recommendations
Areerakulkan et al. (2024)	Holt-Winters, ARIMA, ETS, ANN	Mobile phone sales and inventory data	ANN achieved 51.28% higher accuracy compared to traditional statistical models	Reduces loss sales and inventory carrying costs by 27.71%	Needs evaluation on broader datasets with real-time data; further enhancement with hybrid models
Lei et al. (2024)	Transfer learning framework integrating different aggregation levels	Top-level and bottom-level sales data from JD.com and Walmart	9% improvement in forecasting performance compared to JD.com's benchmark method	Generalizability validated on Walmart data, significant cost savings for low-margin businesses	Needs exploration of alternative deep learning techniques for better accuracy; scalability analysis for different product categories
Yani and Aamer (2023)	Machine learning (Random Forest, Simple Tree)	Pharmaceutical demand and supply chain data	Improvement in demand forecasting accuracy from 10% to 41%	Outperformed other models in pharmaceutical domain	Requires integration with real-world supply chain management systems for practical applicability
Shibu and Agarwal (2023)	Seasonal Naïve, Holt-Winters Triple Exponential Smoothing, Seasonal ARIMA, Linear Regression	Five years of historical sales data	Error rate did not exceed 27% across models	Determines impact of different models on business performance improvement	Needs incorporation of deep learning approaches for better accuracy and adaptability
Ma et al. (2023)	SARIMA-LSTM-BP combination model	New energy vehicle sales and demand data	Lower RMSE, MSE, and MAE compared to traditional models, demonstrating higher forecasting accuracy	Outperforms single-model approaches like RF, SVR, LSTM, BP	Needs testing on additional datasets and real-time implementation for dynamic forecasting
Taghiyeh et al. (2023)	Multi-phase hierarchical approach using machine learning	Hierarchical sales data from a logistics solutions provider	82–90% improvement in forecast accuracy by leveraging child-level forecasts to enhance parent-level predictions	Reduces logistics costs significantly, especially in e-commerce	Further validation needed on different industry datasets; potential integration with real-time demand sensing

III. METHODOLOGY

The proposed methodology for demand forecasting using machine learning involves structured steps to process Walmart data for accurate sales prediction. Data cleaning, one-hot encoding, and normalization are some of the preprocessing procedures used to guarantee high-quality data at the beginning. Feature selection is then performed to identify the most relevant attributes for training models. The dataset is split into 80% training and 20% testing subsets to enable robust model evaluation. For predictive modeling, the study implements XGBoost and LightGBM, comparing their performance with RF and KNN. Models are evaluated using key metrics, including R^2, RMSE, MAE, and MSE. As a last step, the data is evaluated to find the best demand forecasting model. Figure 1 shows the methods that will help enhance the accuracy and efficiency of sales trend predictions, which will help strengthen supply chains.

➢ *Data Collection*

The Walmart dataset holds historical sales data about different Walmart stores, along with their departments and weekly sales records. The dataset consists of multiple variables that include Store ID, Department, Date, Weekly Sales, Holiday Flag, Temperature, Fuel Price, CPI (Consumer Price Index), and Unemployment (the unemployment rate), together with additional economic indicators.

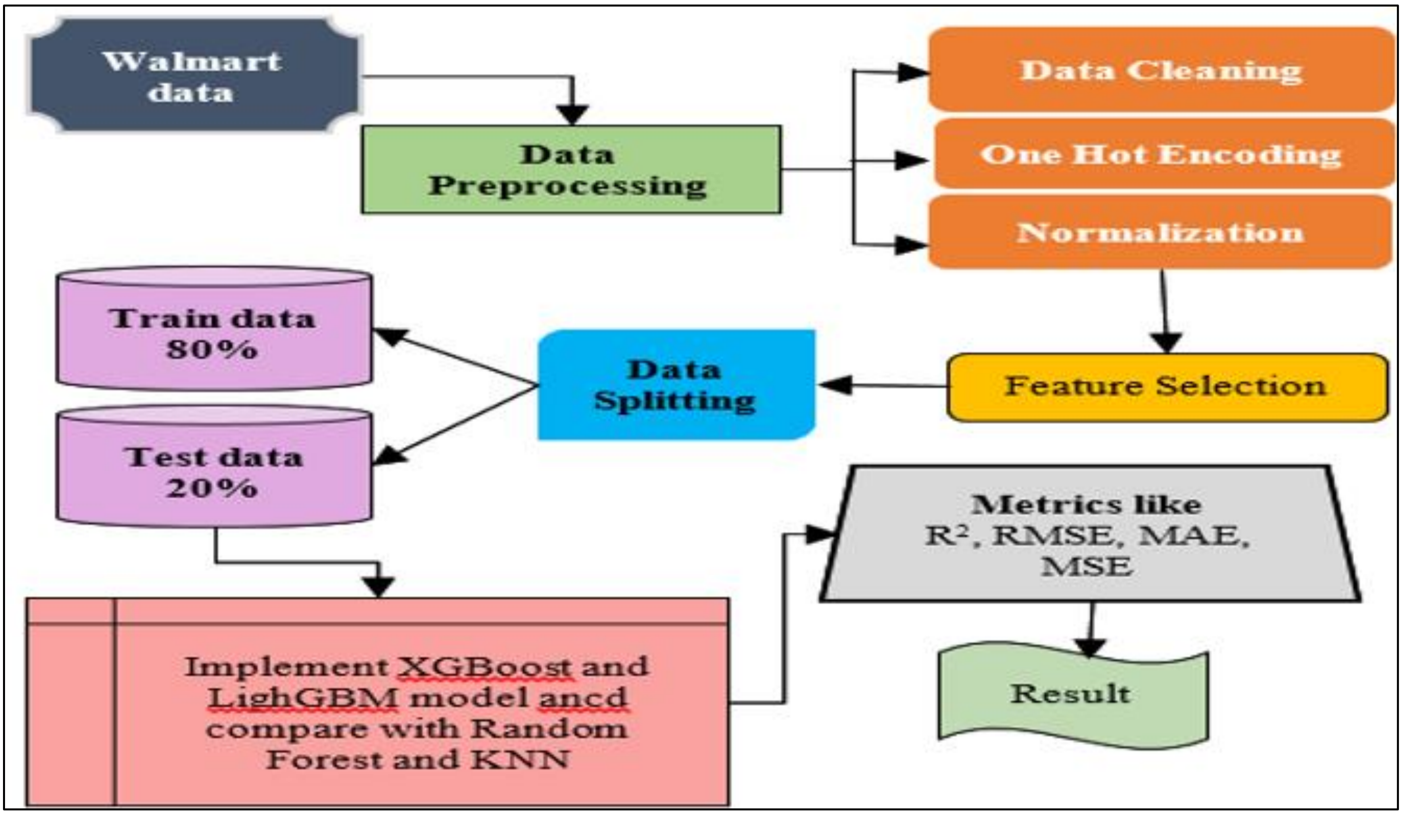

Fig 1 Workflow Of A1-Based Demand Forecasting For Supply Chain Using Walmart Data

The following whole process of proposed methodology are discussed below:

- *Exploratory Data Analysis (EDA)*

An essential first step in comprehending the dataset before implementing ML models is exploratory data analysis, or EDA. This section involves visualizations of the data.

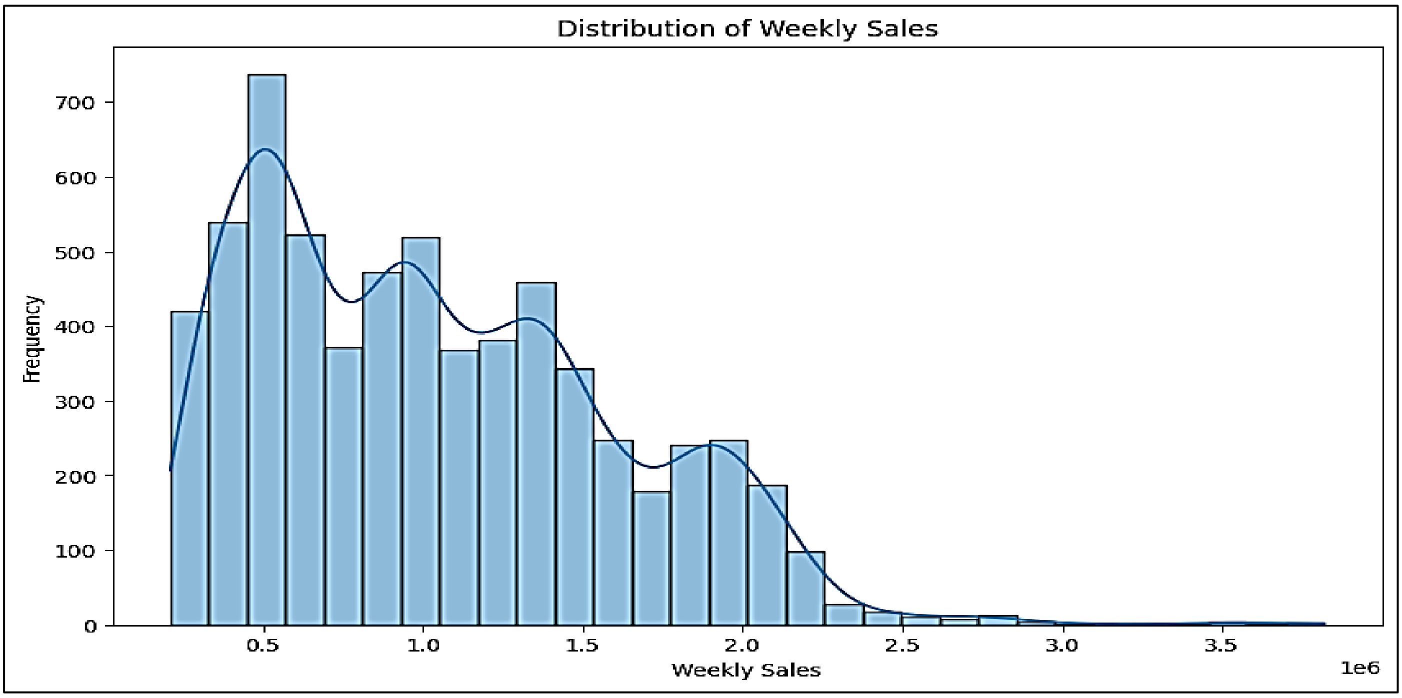

Fig 2 Weekly Sales Distribution In Walmart Stores

Figure 2 illustrates the distribution of Walmart's weekly sales, with the x-axis representing sales values and the y-axis indicating frequency. The data is right-skewed, showing most weekly sales concentrated below $2 million. A KDE curve overlays the histogram, highlighting peaks and variations in sales trends.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr2260

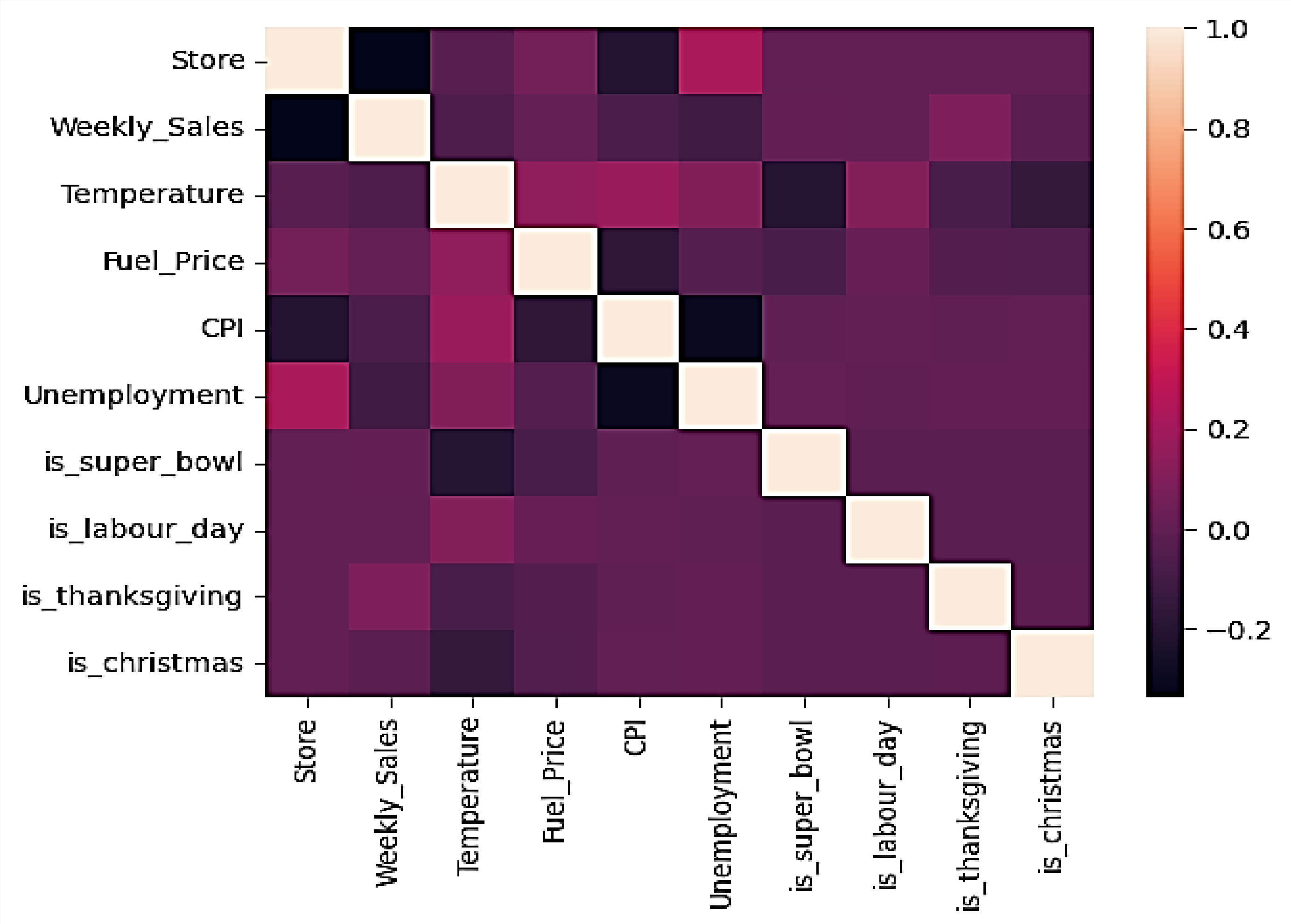

Fig 3 Correlation Matrix of Walmart Dataset

The heatmap in Figure 3 visualizes the correlation among various features in Walmart's sales dataset. Darker shades indicate weaker or negative correlations, while lighter shades represent stronger relationships. Weekly sales exhibit correlations with factors like fuel prices, unemployment, and holidays, providing insights into how external conditions influence Walmart's sales performance.

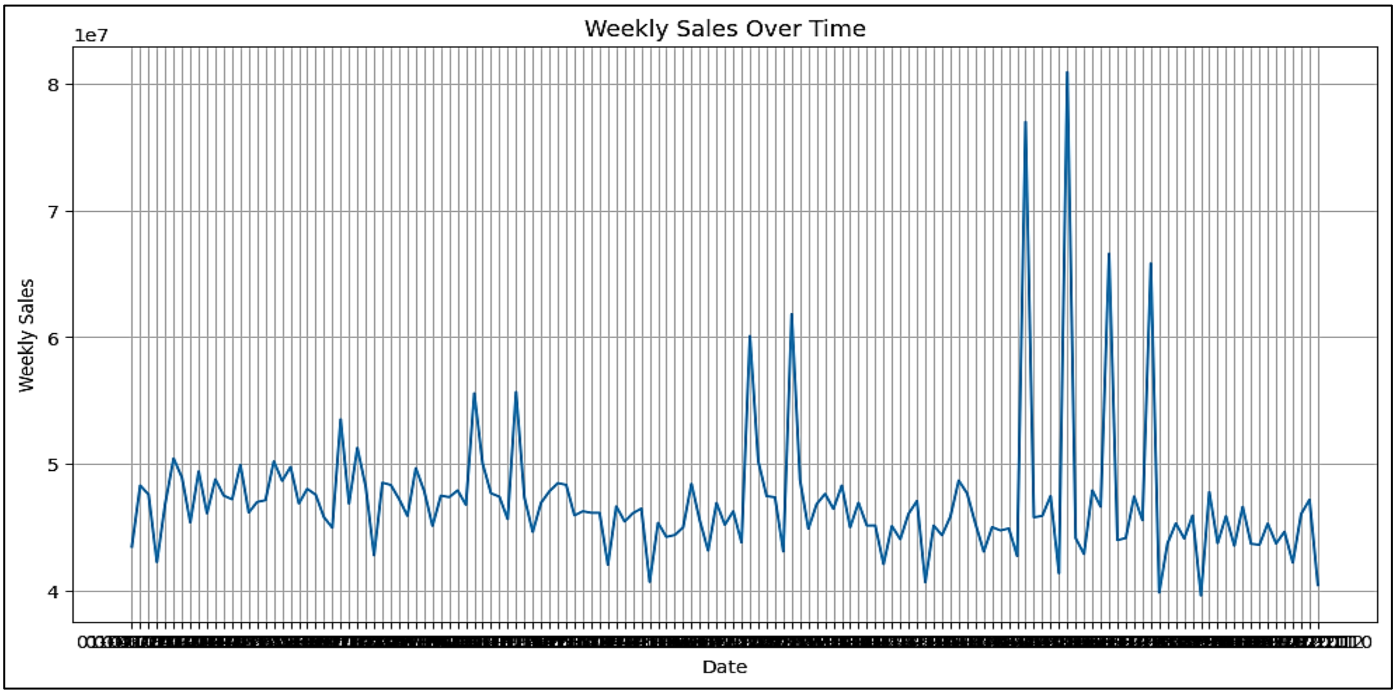

Fig 4 Weekly Sales Trends in Walmart (2010-2012)

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr2260

Figure 4 illustrates Walmart's weekly sales trends over time, with the x-axis representing dates and the y-axis depicting sales figures. Sales generally fluctuate between $40 million and $50 million, with occasional sharp spikes exceeding $70 million, likely due to seasonal events, holidays, or promotional activities.

➢ *Data Preprocessing*

Data Preprocessing turns raw data suitable for machine learning models, including cleaning it before performing necessary transformations and proper organization [23]. The preprocessing performed in this study has the following steps:

- Instances or null values are removed to tidy up the dataset [24]. Removable instances are those that include any inaccessible values.
- There are no duplicate values in the dataset since there are about the same number of rows as there are unique values in the columns [25]. However, if duplicate records were present, removing them would be necessary to ensure data integrity and prevent redundancy, which could otherwise skew analysis and model performance.

➢ *One-Hot Encoding (Binary Encoding)*

One-hot encoding is a method employed to transform category input into numerical values comprehensible to machine learning models. The Holiday_Flag feature was subdivided into four distinct binary features: is_super_bowl, is_labour_day, is_thanksgiving, and is_christmas. Each of these additional columns indicates whether a specific date matches a particular holiday (1) or not (0). This transformation enables models to differentiate between many holidays rather than categorizing them as a singular occurrence. Encoding holidays independently enables the model to discern the distinct impact of each holiday on sales patterns and trends, hence enhancing prediction accuracy.

➢ *Normalization With Min Max Scaler*

The purpose of the normalization phase is to provide a more suitable range for the values of an attribute [26]. Additional methods exist for data normalization. This research choose to scale their data using a technique that is comparable to the min-max approach, which linearly transforms data within a specified interval. The min-max scale Equation is shown in (1).

$$X' = \frac{X - minX}{max - minX} \tag{1}$$

➢ *Feature Selection Using Selectkbest Method*

A utility from the Scikit-learn toolkit, the SelectKBest feature selection technique eliminates all features save the ones with the highest scores. the Holiday_Flag and Date columns were dropped because they did not contribute significantly to the target variable. SelectKBest ranked features based on their correlation with Weekly_Sales, ensuring that only the most impactful ones were retained. This process reduces model complexity, decreases training time, and prevents overfitting. By automatically selecting the most meaningful columns, the dataset becomes more efficient, allowing the model to focus on relevant information and improve overall predictive performance while eliminating less informative variables.

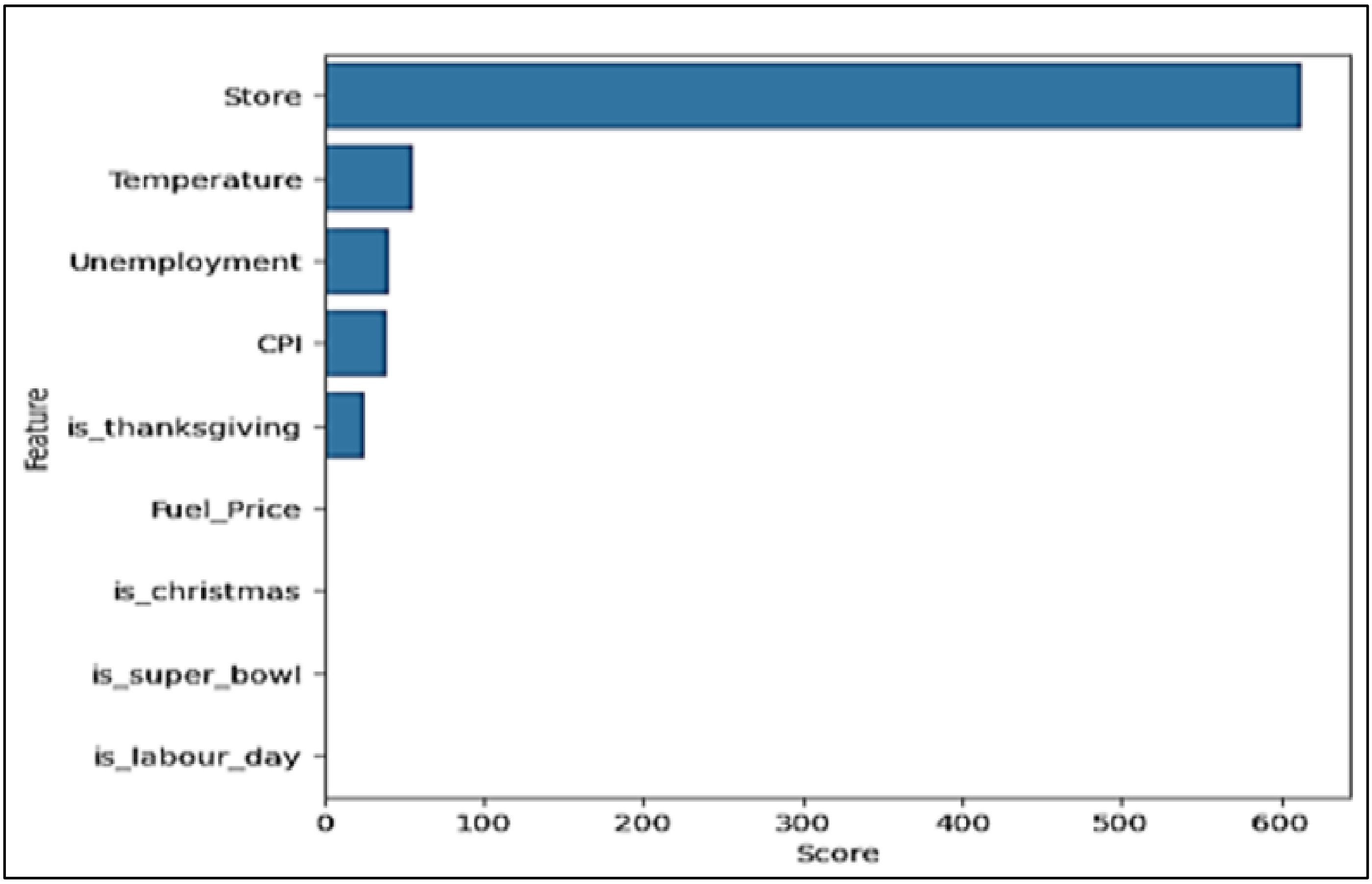

Fig 5 Feature Importance in Walmart Sales Prediction

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr2260

Figure 5 shows feature importance for Walmart data, with "Store" having the highest impact. "Temperature," "Unemployment," and "CPI" have moderate influence, while holidays and fuel prices contribute minimally. This analysis helps prioritize key features for improving Walmart's sales forecasting models.

- *Data Splitting*

Data splitting is the process of splitting a dataset into training and testing subsets to assess model performance. An 80:20 split uses 80% of the data for training, ensuring that the model learns patterns, and 20% for testing, evaluating generalization and preventing overfitting.

- *Implementation of Predication Using XGBoost*

The XGBoost method is a machine learning technique that uses a series of low-quality predictions. A refinement of the gradient boosting technique, this method was first proposed by Chen. Trees are estimated iteratively using the residuals produced at each step of gradient boosting, and the estimates are adaptively updated [27]. The gradient descent method is used in gradient boosting, and the split that supports the approach to the objective function's minimum point is selected. XGBoost's advantages over gradient boosting include its scalability, use of distributed and parallel computation, tree pruning, handling of missing information, and regularization to prevent bias and overfitting.

A popular tree-based approach called XGBoost generates trees one after the other, fixing the mistakes of the previous one. Its objective function is displayed in the Equation (2):

$$L(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \Upsilon||\theta||^1 + \lambda||\theta||^2 \tag{2}$$

- *Implementation of Predication Using LightGBM*

As foundational learners, decision trees power LightGBM, a gradient-boosting architecture. Big datasets become manageable because of the efficiency features that this system offers. In order for LightGBM to work, decision trees are incrementally added with the purpose of fixing errors committed by earlier trees. It improves performance and memory utilization by selecting the most informative cases for tree construction using gradient-based one-sided sampling. For regression tasks including review data, LightGBM may be used to forecast numerical quantities like review count or star rating. The gradient boosting framework of LightGBM adds decision tree leaves successively to build its model. Its objective is to minimize the loss function presented in the Equation (3):

$$L(\theta) = \sum_{i=1}^{n} l\big(y_i, \hat{y}_i(\theta)\big) + \lambda||\theta||^1 \tag{3}$$

where $y_i, \hat{y}_i(\theta)$ is the loss function (e.g., mean squared error), λ is the regularization term, and θ represents model parameters. Parameters are updated using gradient descent displayed in the Equation (4):

$$\theta^{(t+1)} = \theta^{(t)} - \alpha \nabla_\theta L(\theta) \tag{4}$$

Where α is the learning rate.

- *Performance Metrics for Evaluation*

The results of regression models are assessed by different quality characteristics, which define the models' accuracy, stability, and ability to generalize. These measures assist in evaluating the accuracy of a model regarding the continuous numerical values and its proficiency to detect these patterns. Key metrics include:

- *R^2 (R-squared)*

The square of the correlation among the actual and predicted variables, or the percentage of the forecasted variable explained by the regression model, is represented as R2. The formula is shown in Equation (5).

$$R^2 = 1 - \frac{\Sigma(y_i - \hat{y})^2}{\Sigma(y_i - \bar{y})^2} \tag{5}$$

- *RMSE (Root Mean Squared Error)*

The RMSE measures the extent to which actual values deviate from predictions. The formula is presented in Equation (6).

$$RMSE = \sqrt{\frac{1}{n} * \sum_{x=1}^{n} |(y(x) - y'(x)|^2} \tag{6}$$

- *MAE*

The Mean Absolute Error shows how far off the predicted and measured values are on average. The accuracy of the forecast increases when the MAE metric gets closer to zero. The Equation (7) presented the formula of MAE.

$$MAE = \sum_{t=1}^{n} \frac{|Y_t - \hat{Y}_t|}{n} \tag{7}$$

- *MSE*

The Mean Squared Error computation relies on Equation (8) using actual load values (y) and predicted values (y˘) along with n representing the sample number.

$$MSE = \frac{1}{n} \sum_{1}^{n} (Y - \hat{Y})^2 \tag{8}$$

These evaluation measures assess the model's output on test data in order to evaluate its practical application value.

IV. RESULT ANALYSIS AND DISCUSSION

The research findings of AI-based demand projection with XGBoost and LightGBM are presented in this section. A state-of-the-art computer setting equipped with an NVIDIA GPU for fast training alongside 32 GB of RAM and an Intel Core i7 CPU conducted the experimental runs. The models built their structure in Jupyter through Python modules Scikit-learn, XGBoost and LightGBM. The predictive accuracy metrics included R², RMSE, MAE and MSE for evaluation. All test data results demonstrate that LightGBM manages big demand datasets well, but XGBoost leads to better capture of demand shifts, which strengthens supply chain stability.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr2260

Table 2 Xgboost and Lightgbm Performance for Demand Forecasting Using Artificial Intelligence

Metrics	XGBoost	LightGBM
R-square	0.9752	0.9732
MSE	0.008826	0.0095514
RMSE	0.09394	0.09773
MAE	0.06188	0.06471

Table 2 displayed the performance of XGBoost and LightGBM models for demand forecasting using Artificial Intelligence. The tool presents four evaluation metrics including R-squared together with MSE, RMSE and MAE. The R-squared value in both models delivers close to 1 indicating strong data fit. MSE, RMSE, and MAE are low for both, suggesting accurate predictions. XGBoost and LightGBM exhibit similar performance across all metrics, with slight variations. The graph highlights the effectiveness of these AI techniques in demand forecasting.

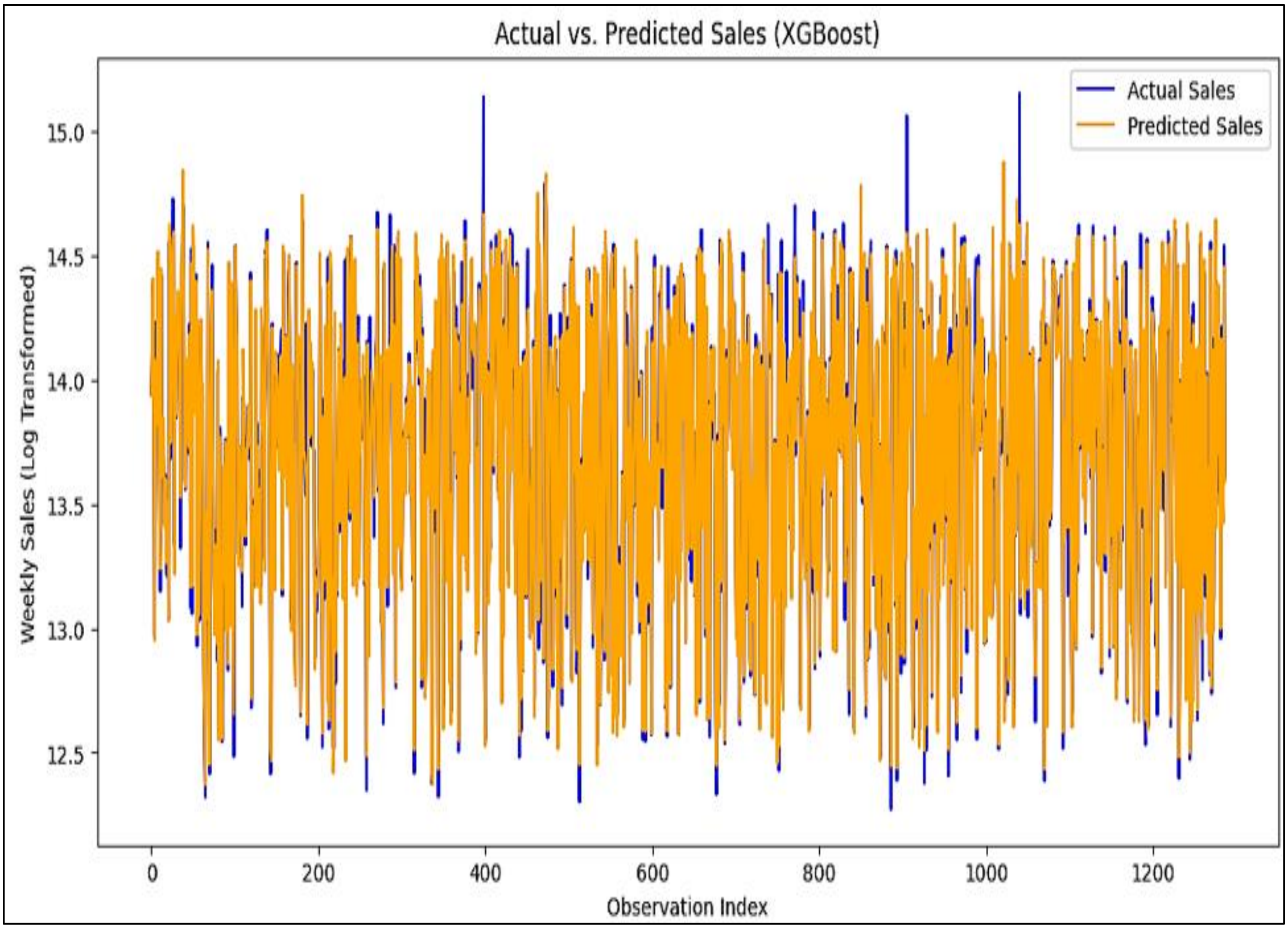

Fig 6 Actual Vs. Predicted Weekly Sales Using Xgboost For Demand Forecasting

Figure 6 is a line plot comparing actual and predicted sales using XGBoost. The x-axis represents the observation index, while the y-axis shows log-transformed weekly sales. Blue lines indicate actual sales, and orange lines represent predictions. The dense overlap suggests strong model performance. A legend in the top-right labels both lines, highlighting the model’s accuracy in forecasting sales trends.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr2260

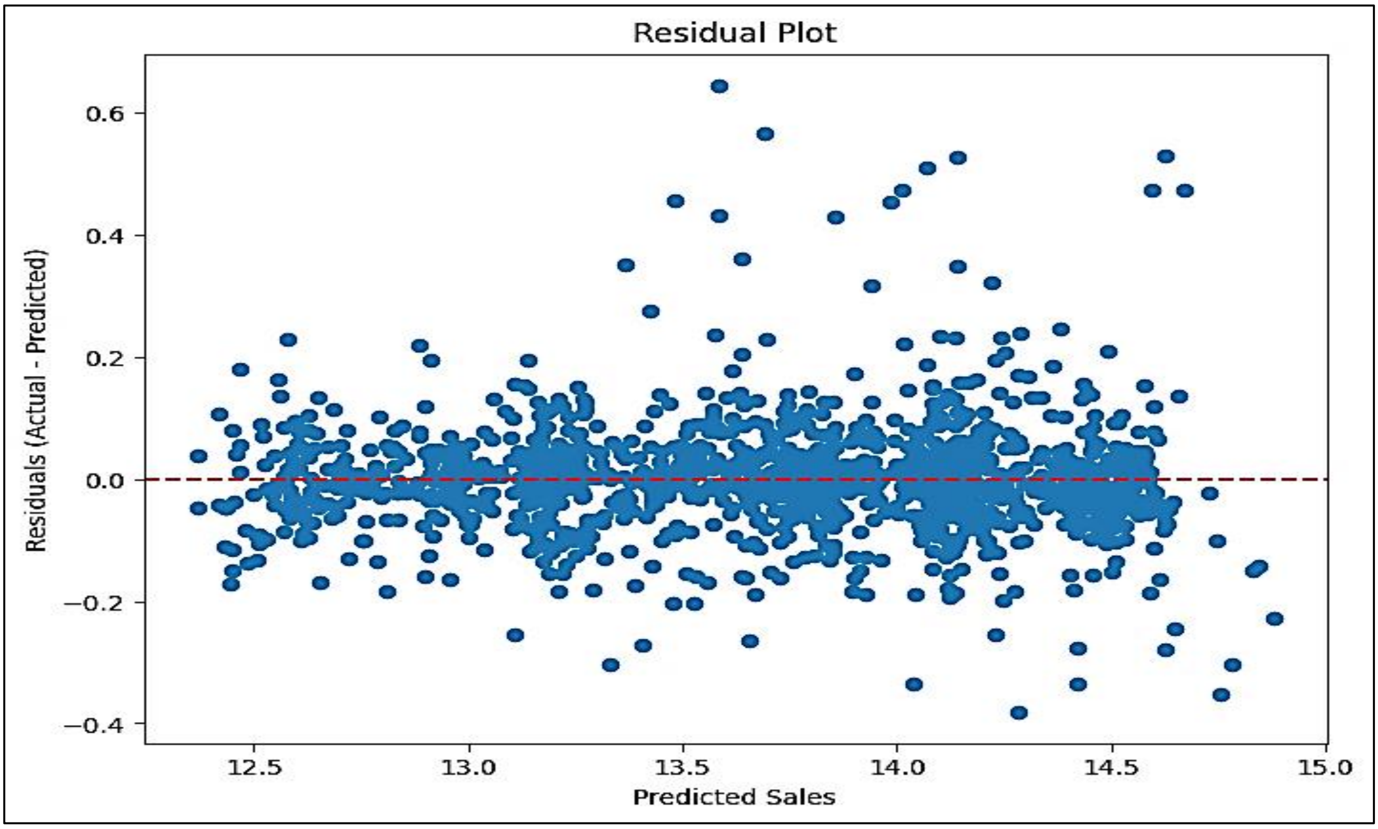

Fig 7 Residual Plot for Xgboost Model in Demand Forecasting

The residual plot in Figure 7 displays the differences between actual and predicted sales. The x-axis represents predicted sales, while the y-axis shows residuals (Actual - Predicted). Blue dots indicate residual values, and a red dashed line represents the zero-error baseline. The residuals are scattered around zero without a clear pattern, suggesting no major heteroscedasticity issues and that the model performs well without significant bias across different sales values.

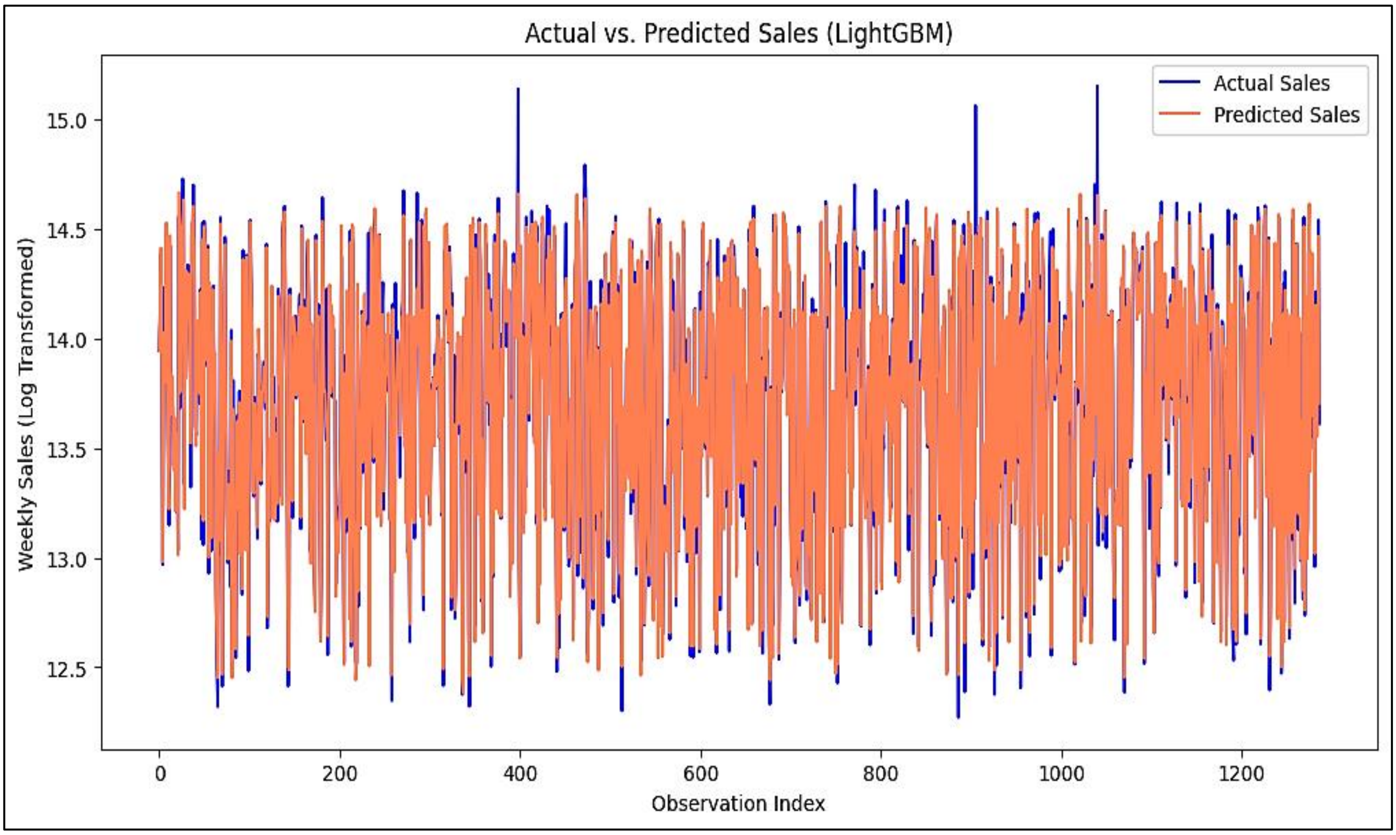

Fig 8 Actual Vs. Predicted Weekly Sales Using Lightgbm For Demand Forecasting

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr2260

A line plot comparing actual and expected sales employing the LightGBM model is shown in Figure 8. On one side, it has the observation index, and on the other, it has the log-transformed weekly sales. Forecasted sales are shown by the orange line, while actual sales are shown by the blue line. The close alignment of both lines suggests that the model performs well. A legend in the top-right corner labels the actual and predicted sales.

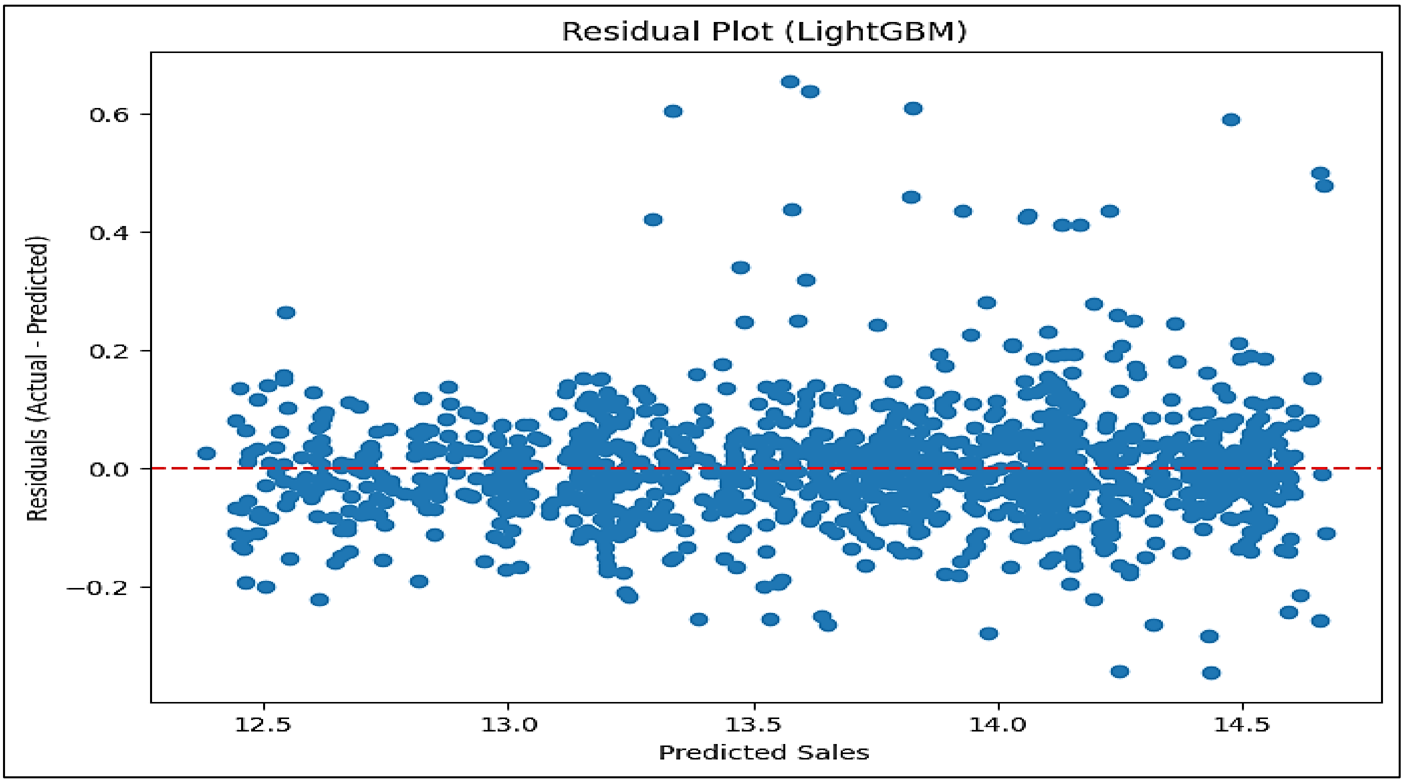

Fig 9 Residual Plot for LightGBM Model in Demand Forecasting

Figure 9 shows the disparity between the actual and expected sales, as shown in the LightGBM model's residual plot. The x-axis represents predicted sales, while the y-axis displays residuals (Actual - Predicted). Blue dots represent residual values, and the red dashed line marks the zero-error baseline. Residuals are scattered around zero without a clear pattern, indicating minimal bias and suggesting that the model captures the data well.

➢ *Comparison and Discussion*

Table III present a comparative analysis of XGBoost, LightGBM, RF, and K-Nearest Neighbors (KNN) for AI-driven demand forecasting. XGBoost, a gradient-boosting algorithm, demonstrates superior performance with an R^2 score of 0.9752 and the lowest MSE of 0.008826, making it the most effective model for demand prediction. LightGBM, optimized for large-scale datasets, achieves an R^2 score of 0.9732 with an MSE of 0.0095514, showcasing its efficiency in handling complex demand variations. Random Forest and KNN exhibit lower predictive accuracy, with R^2 scores of 0.9569 and 0.9381, respectively, and significantly higher MSE values, indicating their limitations in capturing intricate demand patterns.

Table 3 Comparative Performance of ML Models for Demand Forecasting

Models	R^2	MSE
XGBoost	0.9752	0.008826
LightGBM	0.9732	0.0095514
Random Forest[28]	0.9569	2.243
KNN[29]	0.93810	1.994

The proposed approach offers several advantages in AI-driven demand forecasting for supply chain resilience. By leveraging XGBoost and LightGBM, the models achieve high predictive accuracy, with XGBoost attaining an R^2 score of 0.9752 and the lowest MSE of 0.008826. Automated feature selection enhances efficiency, reducing reliance on data processing, improve data quality, ensuring robust performance. The models generalize well with minimal overfitting, enabling precise demand forecasting. This enhances inventory optimization, reduces stock fluctuations, and supports proactive decision-making in supply chain management.

V. CONCLUSION AND FUTURE SCOPE

AI-driven demand forecasting has proven to be a transformative approach for enhancing supply chain resilience. Demand forecasting is crucial for retail businesses like Walmart to optimize inventory management and enhance profitability. This study demonstrates the effectiveness of

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr2260

machine learning models, particularly XGBoost and LightGBM, in demand forecasting for Walmart sales data. Through structured data preprocessing, feature selection, and model evaluation, both models exhibit high predictive accuracy, with R^2 values of 0.9752 and 0.9732, respectively. Compared to Random Forest ($R^2 = 0.9569$) and KNN ($R^2 = 0.9381$), XGBoost and LightGBM provide superior performance with lower MSE, RMSE, and MAE, indicating their robustness in handling sales data fluctuations. The residual analysis confirms minimal bias and no significant heteroscedasticity issues, further validating the models' reliability. Overall, this research highlights the potential of advanced gradient boosting techniques for accurate demand forecasting, which can aid retail businesses in optimizing inventory management, reducing stock shortages, and improving sales strategies. Despite the high accuracy of XGBoost and LightGBM, the study has limitations, including sensitivity to hyperparameter tuning and potential overfitting to specific trends. Additionally, the models do not account for external factors such as economic conditions, holidays, and promotions, which can impact sales. Future research can focus on integrating real-time forecasting capabilities using deep learning models like LSTMs and transformers to capture temporal dependencies. Additionally, incorporating external factors such as economic indicators, promotions, and weather conditions may further refine demand predictions. Expanding this approach to multi-channel retailing and supply chain disruptions could provide deeper insights into demand variability, ensuring greater adaptability and efficiency in dynamic market environments.

REFERENCES

[1]. M. Abolghasemi, E. Beh, G. Tarr, and R. Gerlach, “Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion,” Comput. Ind. Eng., 2020, doi: 10.1016/j.cie.2020.106380.

[2]. M. Seyedan and F. Mafakheri, “Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities,” J. Big Data, 2020, doi: 10.1186/s40537-020-00329-2.

[3]. K. Murugandi and R. Seetharaman, “Analysing the Role of Inventory and Warehouse Management in Supply Chain Agility: Insights from Retail and Manufacturing Industries,” Int. J. Curr. Eng. Technol., vol. 12, no. 6, pp. 583–590, 2022.

[4]. H. S. Chandu, “Enhancing Manufacturing Efficiency: Predictive Maintenance Models Utilizing IoT Sensor Data,” IJSART, vol. 10, no. 9, 2024.

[5]. M. A. Khan et al., “Effective Demand Forecasting Model Using Business Intelligence Empowered with Machine Learning,” IEEE Access, 2020, doi: 10.1109/ACCESS.2020.3003790.

[6]. Suhag Pandya, “Innovative blockchain solutions for enhanced security and verifiability of academic credentials,” Int. J. Sci. Res. Arch., vol. 6, no. 1, pp. 347–357, Jun. 2022, doi: 10.30574/ijsra.2022.6.1.0225.

[7]. S. Chatterjee, “Mitigating Supply Chain Malware Risks in Operational Technology: Challenges and Solutions for the Oil and Gas Industry,” J. Adv. Dev. Res., vol. 12, no. 2, pp. 1–12, 2021.

[8]. F. Z. Benhamida, O. Kaddouri, T. Ouhrouche, M. Benaichouche, D. Casado-Mansilla, and D. López-De-Ipiña, “Demand forecasting tool for inventory control smart systems,” J. Commun. Softw. Syst., 2021, doi: 10.24138/jcomss-2021-0068.

[9]. K. M. R. Seetharaman, “End-to-End SAP Implementation in Global Supply Chains: Bridging Functional and Technical Aspects of EDI Integration,” Int. J. Res. Anal. Rev., vol. 8, no. 2, pp. 1–7, 2021.

[10]. S. Murri, “Optimising Data Modeling Approaches for Scalable Data Warehousing Systems,” Int. J. Sci. Res. Sci. Eng. Technol., vol. 10, no. 5, pp. 369–382, Oct. 2023, doi: 10.32628/IJSRSET2358716.

[11]. E. Hofmann and E. Rutschmann, “Big data analytics and demand forecasting in supply chains: a conceptual analysis,” Int. J. Logist. Manag., vol. 29, no. 2, pp. 739–766, 2018.

[12]. S. Tyagi, “Analyzing Machine Learning Models for Credit Scoring with Explainable AI and Optimizing Investment Decisions,” Am. Int. J. Bus. Manag., vol. 5, no. 01, pp. 5–19, 2022.

[13]. S. Tyagi, T. Jindal, S. H. Krishna, S. M. Hassen, S. K. Shukla, and C. Kaur, “Comparative Analysis of Artificial Intelligence and its Powered Technologies Applications in the Finance Sector,” in Proceedings of 5th International Conference on Contemporary Computing and Informatics, IC3I 2022, 2022. doi: 10.1109/IC3I56241.2022.10073077.

[14]. S. Arora and S. R. Thota, “Using Artificial Intelligence with Big Data Analytics for Targeted Marketing Campaigns,” Int. J. Adv. Res. Sci. Commun. Technol., vol. 4, no. 3, pp. 593–602, Jun. 2024, doi: 10.48175/IJARSCT-18967.

[15]. Olamide Raimat Amosu, Praveen Kumar, Yewande Mariam Ogunsuji, Segun Oni, and Oladapo Faworaja, “AI-driven demand forecasting: Enhancing inventory management and customer satisfaction,” World J. Adv. Res. Rev., vol. 23, no. 2, pp. 708–719, 2024, doi: 10.30574/wjarr.2024.23.2.2394.

[16]. J. Feizabadi, “Machine learning demand forecasting and supply chain performance,” Int. J. Logist. Res. Appl., 2022, doi: 10.1080/13675567.2020.1803246.

[17]. N. Areerakulkan, C. Moryadee, L. Trakoonsanti, M. Khaengkhan, and N. Setthachotsombut, “A New Product’s Demand Forecasting Using Artificial Neural Network,” Int. Conf. Enterp. Inf. Syst. ICEIS - Proc., vol. 1, no. Iceis, pp. 417–424, 2024, doi: 10.5220/0012734800003690.

[18]. D. Lei et al., “Pooling and boosting for demand prediction in retail: A transfer learning approach,” Manuf. \& Serv. Oper. Manag., 2024.

[19]. L. P. E. Yani and A. Aamer, “Demand forecasting accuracy in the pharmaceutical supply chain: a machine learning approach,” Int. J. Pharm. Healthc. Mark., vol. 17, no. 1, pp. 1–23, 2023.

[20]. N. Shibu and R. Agarwal, “Analysing and Visualising trends for Supply Chain Demand Forecasting,” in Proceedings of International Conference on Computational Intelligence and Sustainable Engineering Solution, CISES 2023, 2023. doi: 10.1109/CISES58720.2023.10183439.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr2260

[21]. X. Ma, M. Li, J. Tong, and X. Feng, "Deep Learning Combinatorial Models for Intelligent Supply Chain Demand Forecasting," Biomimetics, vol. 8, no. 3, 2023, doi: 10.3390/biomimetics8030312.

[22]. S. Taghiyeh, D. C. Lengacher, A. H. Sadeghi, A. Sahebi-Fakhrabad, and R. B. Handfield, "A novel multi-phase hierarchical forecasting approach with machine learning in supply chain management," Supply Chain Anal., 2023, doi: 10.1016/j.sca.2023.100032.

[23]. B. Boddu, "Scaling Data Processing with Amazon Redshift Dba Best Practices for Heavy Loads," Int. J. Core Eng. Manag., vol. 7, no. 7, 2023.

[24]. A. Alabrah, "An Improved CCF Detector to Handle the Problem of Class Imbalance with Outlier Normalization Using IQR Method," Sensors, vol. 23, no. 9, 2023, doi: 10.3390/s23094406.

[25]. K. M. Alalayah, E. M. Senan, H. F. Atlam, I. A. Ahmed, and H. S. A. Shatnawi, "Automatic and Early Detection of Parkinson's Disease by Analyzing Acoustic Signals Using Classification Algorithms Based on Recursive Feature Elimination Method," Diagnostics, 2023, doi: 10.3390/diagnostics13111924.

[26]. A. Băicoianu, C. G. Gavrilă, C. M. Păcurar, and V. D. Păcurar, "Fractal interpolation in the context of prediction accuracy optimization," Eng. Appl. Artif. Intell., vol. 133, p. 108380, 2024, doi: https://doi.org/10.1016/j.engappai.2024.108380.

[27]. A. Massaro, A. Panarese, D. Giannone, and A. Galiano, "Augmented Data and XGBoost Improvement for Sales Forecasting in the Large-Scale Retail Sector," Appl. Sci., vol. 11, no. 17, 2021, doi: 10.3390/app11177793.

[28]. C. Yu, "Walmart Sales Forecasting using Different Models," Highlights Sci. Eng. Technol., vol. 92, pp. 302–307, 2024, doi: 10.54097/kqf76062.

[29]. A. Tayyab, "Walmart Sales Prediction By Using Machine Learning Algorithms," J. Innov. Comput. Emerg. Technol., vol. 4, 2024, doi: 10.56536/jicet.v4i2.132.