IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14410

Interpretable vs Predictive: Comparing Statistical and ML Models of Walmart Black Friday Demographics in Consumer Behaviour Analysis

Harish I[1], Arun Kumar K[2]

Department of M.Sc. Data Science and Business Analysis, Rathinam College of Arts and Science, Coimbatore[1]

Department of Computer Science, Rathinam College of Arts and Science, Coimbatore[2]

Abstract: Walmart's Black Friday, falling the day after Thanksgiving in the U.S., marks the launch of holiday shopping and drives massive retail sales nationwide. Retailers face the challenge of balancing interpretable statistical insights with predictive machine learning to analyze consumer behavior. This research proposes a unified framework that emphasizes clarity, actionable insights, and predictive strength to guide retail decision-making. By integrating statistical methods with machine learning, the study enhances both retail insights and forecasting power. Looking ahead, retail analytics is poised to evolve through deep learning, real-time processing, and scalable data platforms, enabling faster adaptation to shifting consumer trends and competitive dynamics.

Keywords: Black Friday - Consumer Behavior - Machine Learning - statistics -Unified Framework

I.INTRODUCTION

Walmart's Black Friday event is one of America's biggest retail days of the year, occurring the day after Thanksgiving Day. It offers massive price reductions on a wide variety of goods, such as electronics, clothing, household items, and toys, to millions of consumers in both physical and digital stores. In addition to being a retail phenomenon, it is also an opportunity for researchers to investigate consumer behaviour, including consumer purchasing patterns and differences among demographic groups, during a short and very intense buying experience. For businesses, it demonstrates the necessity of using predictive analytics to anticipate demand, manage inventory, and provide customers with personalized offers. Walmart's Black Friday events are both a shopping fiesta and an example of retail analytics applied in a short-term period. Analysing retail sales using analytics tools is a critical part of a company's business strategy today, especially for companies like Walmart who engage in large-scale retail operations during events like Black Friday by drawing on insights keeping the same tone while ensuring from consumer behaviour analysis, companies can refine their marketing strategies, streamline inventory pricing models, and strengthen customer management, adjust engagement. The primary challenge is selecting an appropriate analytical strategy. That is, should researchers utilize interpretable statistical models (which are transparent), or predictive machine learning models (which prioritise accuracy)? Traditional statistical analyses- including confidence intervals, hypothesis tests, ANOVA/MANOVA, and logistic regression have been recognised for their interpr`etability for many years. Statistical techniques can be applied to confirm sample reliability, explore demographic differences, and estimate the probability of higher spending based on customer characteristics through structured models. Traditional statistical approaches are useful in determining the justification for spending differentially among various demographic groups such as age group, gender, marital status, income level, occupation, and geographical area (where someone lives). Conversely as K-Means clustering, machine learning techniques such, decision trees, and XGBoost give, random forests researchers strong predictive power to uncover complex and the ability with greater precision customer segments. The machine learning models demonstrate high levels of prediction accuracy as well as the ability to identify segments of customers who were difficult for traditional statistical methods to identify. Additionally, machine learning models can generate rule-based predictions and quantify feature importance with high accuracy; whereas traditional statistical models place less emphasis on explaining purchasing behaviour and instead place greater emphasis on forecasting what will happen in the future as it relates to predicting spending behaviour. The dataset for this research is comprised of a combination of demographic variables (i.e., ethnicity, age, income, education level) and purchase levels associated with specific products. product-specific. The dataset, which combines demographic information with lens through which purchase records, offers a comprehensive consumer decision across diverse factors-making can be understood. Furthermore, it allows analysts to detect nuanced behavioural patterns that traditional statistical approaches often overlook, thereby enriching the overall interpretation. Moreover, it enables researchers to uncover subtle behavioral trends that remain hidden in traditional analyses. Customer_ID makes each customer unique; while demographics (or "demographics") and socioeconomic attributes give each customer a rich set of characteristics to

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14410

be interpreted as they relate to spending patterns; Product_ID and Product_Category provide the product type for the segmentation; Purchase_Amount is the dependent variable used for making both inferences and predictions. The research prefaces interpretability and prediction not as mutually exclusive approaches, but rather as complementary tools to be used together. The statistical model identifies which variables (or "drivers") are important to consumers' spending patterns, while the machine learning models provide better forecasts and better segmented product offerings. These two complementary approaches yield a combined framework to address the tension between the need for businesses to interpret spending patterns and the need for businesses to have the predicted level of accuracy of forecasting spending patterns. The significance of applying this integrated approach to Walmart's Black Friday database will contribute to both academic research and real-world retail decision-making, as it provides insight into spending behaviour that is both transparent and actionable, as well as having a high level of predictive capability.

II.LITERATURE SURVEY

Kelly, Scott & Bowman (2025) – Influence of Sales Promotion on Consumer Buying Behavior
Sales promotions such as coupons, discounts, free shipping, and BOGO offers strongly influence consumer buying behavior, with the highest overspending seen during Black Friday and Cyber Monday. Economic concepts like price elasticity and marginal utility explain why aggressive discounts drive greater demand.
Bakhrun et al. (2025) – Data Visualization to Analyze Consumer Behavior (Walmart Case Study)
Loyal customers generate most sales, weekday purchases exceed weekends, and Gen X spends the most interactive dashboards turn these insights into stronger marketing, smarter inventory planning, and improved customer retention.
Lennon, Kim, Lee & Johnson (2018) – Consumer Emotions on Black Friday
Black Friday shopping can spark excitement when goals are met, but also frustration when they're missed, showing that emotions directly shape how people judge their experience and make buying decisions.
Neba et al. (2024) – Advancing Retail Predictions with ML (Walmart Sales Forecasting)
Modern machine learning models like Random Forest, GBM, LightGBM, and XGBoost give far more accurate Walmart sales forecasts than old regression methods, with XGBoost showing near-perfect results and helping improve inventory and business decisions.
Patil et al. (2023) – Black Friday Sales Prediction using Supervised ML
XGBoost regression delivers the best results with the lowest error and highest accuracy, while careful data preparation, feature selection, and fair evaluation are key to building strong retail prediction systems.
Vyavahare et al. (2023) – ML Application for Black Friday Sales Prediction Framework
A machine learning framework using features like age, gender, and occupation showed that Random Forest gave strong accuracy and low error in predicting Black Friday spending, proving the value of careful data preparation and fair evaluation for effective retail forecasting+
Ali & Shah (2024) – Predicting Retail Sales for Walmart (ML + Time Series Models)
Hybrid methods that blend machine learning with time series models give higher accuracy than traditional approaches, especially when factors like holidays and weather are included—showing that smart feature design, fair evaluation, and diverse data are key for reliable forecasts and inventory planning.
Elsbree (2022) – Black Friday Pricing Behavior at Walmart
Walmart's Black Friday pricing behavior, tracking over 400 products to evaluate whether discounts are genuine or inflated, and the analysis reveals practices such as reference price inflation and pre-sale price increases, which can make discounts appear larger than they actually are.
Previous research on Walmart's Black Friday sales has mainly focused on boosting prediction accuracy through machine learning models, using metrics such as RMSE, MAE, and R^2. While these studies highlight the strengths of algorithms like Random Forest, Gradient Boosting, and XGBoost, they often overlook economic and statistical perspectives of consumer behavior, including price elasticity, marginal utility, and psychological influences. This gap shows that accurate forecasts alone are not enough without interpretability for strategic retail decisions. To bridge this, the present research introduces a dual-lens framework that combines statistical analysis for clarity with machine learning for predictive strength, producing insights that are academically reliable and practically useful forinventory, pricing, and customer engagement during Black Friday.

III.IMPLEMENT

Workflow of the Project

- Data Collection – Gather demographic details (age, gender, income, occupation, education, location) along with transactional records (product ID, category, purchase amount).

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14410

- Data Preparation – Clean missing values, encode categorical variables, and standardize formats to make the dataset ready for analysis.

- Statistical Analysis – Apply hypothesis testing, ANOVA, and regression to explain spending differences across demographic groups.

- Machine Learning Models – Use Random Forest, Gradient Boosting, LightGBM, and XGBoost to forecast purchase amounts and identify hidden customer segments.

- Integrated Framework – Combine statistical insights (interpretability) with machine learning predictions (accuracy) to form a dual-lens approach.

- Implementation & Insights – Convert findings into strategies for inventory planning, pricing, and personalized marketing during Walmart's Black Friday sales.

Project Scope

This project examines customer spending behavior during Walmart's Black Friday sales, aiming to balance clarity with predictive accuracy. Demographic and transactional data are analysed using statistical methods to explain spending differences, while machine learning models such as Random Forest, Gradient Boosting, LightGBM, and XGBoost are applied for forecasting and segmentation. By integrating these approaches into a dual-lens framework, the project provides insights that support inventory planning, pricing, and personalized marketing.

Collect and Prepare Data

The dataset combines demographic attributes (age, gender, income, education, occupation, location) with transactional details (product ID, category, purchase amount). Each customer is tracked using a unique ID. Preparation includes cleaning missing values, encoding categorical variables, and standardizing formats. This ensures the dataset is suitable for statistical analysis and machine learning, enabling clear interpretation and accurate forecasting.

Statistical Analysis

Statistical methods are used to explain spending behavior with transparency. Techniques such as hypothesis testing, confidence intervals, ANOVA/MANOVA, and logistic regression examine differences across demographic groups. These methods identify key drivers of spending, measure sample reliability, and estimate probabilities of higher purchase amounts. The focus on interpretability explains why groups spend differently and complements machine learning predictions.

1. Confidence Interval (95% CI)

A Confidence Interval estimates the range within which the true population mean is likely to fall. In this project, a 95% CI was calculated for the mean purchase amount across 550,068 transactions. This means if the experiment were repeated 100 times, 95 of those intervals would contain the true average purchase. It helps confirm how reliable the observed mean spending is, rather than treating it as an exact value.

2. Independent Samples T-Test

The T-Test checks whether the average purchase amounts of two groups are statistically different from each other. In this project, two T-Tests were performed:

- Gender T-Test - compares average spending of Male vs. Female customers

- Marital Status T-Test - compares spending of Married vs. Unmarried customers

If the p-value is less than 0.05, the null hypothesis (no difference) is rejected, confirming a statistically significant difference exists between the groups.

3. One-Way ANOVA (Analysis of Variance)

ANOVA extends the T-Test to compare means across three or more groups simultaneously. In this project, ANOVA was applied to five variables: Age groups, Occupation codes, City Category (A/B/C), Years in current city, and Product

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14410

Category. The F-statistic measures how much the group means differ relative to the variation within groups. A significant result means at least one group spends significantly differently from the others.

4. MANOVA (Multivariate Analysis of Variance)

MANOVA is a multi-output extension of ANOVA. While ANOVA tests one dependent variable at a time, MANOVA simultaneously tests multiple outcome variables - here, Purchase amount, Product Category, and Occupation - against demographic factors like Gender, Age, City, and Marital Status. This reveals whether the combined effect of demographics is significant, which individual ANOVAs cannot capture together.

5. Logistic Regression

Logistic Regression is an interpretable statistical model used for binary classification. In this project, it predicts whether a customer is a High Spender (purchase ≥ mean threshold) or Low Spender. It calculates the probability of belonging to each class using a logistic (sigmoid) function. The model's coefficients directly show which features (like Age or Product Category) increase or decrease the likelihood of high spending, making it highly interpretable.

Machine Learning Models

Machine learning improves prediction accuracy and uncovers complex consumer patterns. Models including Random Forest, Gradient Boosting, LightGBM, and XGBoost forecast purchase amounts and segment customers based on demographic and transactional features. These algorithms provide strong predictive power, reveal hidden customer groups, and quantify feature importance. Integrated into the dual-lens framework, they strengthen forecasting and support inventory, pricing, and customer engagement strategies.

1. K-Means Clustering

K-Means is an unsupervised algorithm that groups customers into k clusters based on similarity in their features (Age, Gender, Occupation, Purchase amount, etc.). In this project, k=3 was chosen using the Elbow Method (which plots inertia vs. k and finds the "elbow" where adding more clusters gives diminishing returns). The three clusters segment customers into High Spenders, Medium Spenders, and Low Spenders, enabling targeted marketing for each group.

2. Decision Tree

A Decision Tree splits the data into branches based on feature values (like "Is Age > 26-35?"), forming a tree structure. Each internal node is a decision rule, each branch is an outcome, and each leaf node gives the final prediction. In this project, it was used for both classification (high/low spender) with a max depth of 6, and regression (predict exact purchase amount) with a max depth of 8. The depth limit prevents overfitting. Decision Trees are easy to visualize and explain.

3. Random Forest

Random Forest is an ensemble method that builds 200 decision trees on random subsets of data (a technique called Bootstrap Aggregation / Bagging). Each tree gives a prediction, and the final output is determined by majority vote (classification) or average (regression). Because trees are trained on different data samples and features, they are diverse, which reduces variance and improves accuracy compared to a single Decision Tree. It was used here for both classification and regression.

4. Gradient Boosting Machine (GBM)

Unlike Random Forest (which builds trees independently), Gradient Boosting builds trees sequentially, where each new tree corrects the errors of the previous one. It minimizes a loss function using gradient descent. In this project, 200 trees were built with a learning rate of 0.1 and a max depth of 6. The learning rate controls how much each tree contributes, preventing overfitting. GBM is powerful but slower to train than Random Forest. Used for both classification and regression tasks.

5. LightGBM (Light Gradient Boosting Machine)

LightGBM is an advanced, highly optimized version of Gradient Boosting developed by Microsoft. It uses a leaf-wise tree growth strategy (instead of level-wise), which grows the most impactful leaf first, making it faster and more accurate

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14410

on large datasets. It also uses histogram-based splitting to speed up computation. In this project, 300 estimators were used with a max depth of 8 and learning rate of 0.1. It is especially efficient for the 550,068-row Walmart dataset.

6. XGBoost (Extreme Gradient Boosting)

XGBoost is another high-performance gradient boosting framework that adds regularization (L1 and L2) to the standard boosting process to prevent overfitting. It uses column subsampling (colsample_bytree=0.8) and row subsampling (subsample=0.8) per tree, making it robust and accurate. In this project, 300 trees were built with a max depth of 8. XGBoost was cited by Patil et al. (2023) as the best-performing model for Black Friday purchase prediction, and this project validates that finding.

Integrated Framework

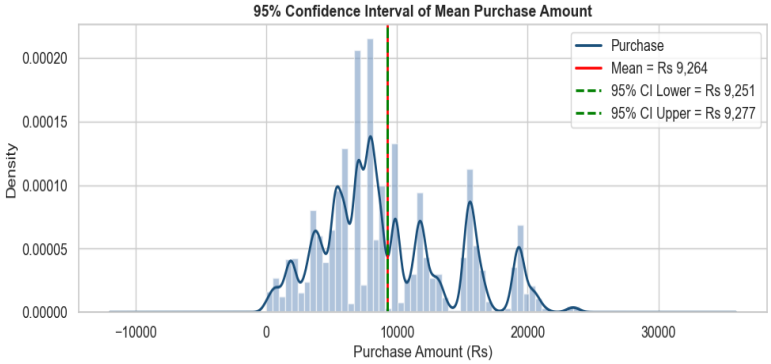
The framework merges statistical analysis and machine learning. Statistical methods highlight demographic and behavioral drivers, while machine learning provides precise forecasts and segmentation. Together, they balance clarity with accuracy, ensuring results are interpretable for decision-makers and actionable for inventory planning, pricing, and marketing during Black Friday.

IV.RESULTS

The analysis of the Walmart Black Friday dataset successfully applied a dual-lens framework, contrasting interpretable statistical methods with powerful machine learning models. The key findings are summarized below.

1. Statistical Analysis (Interpretable Lens)

- **Confidence Interval:** We are 95% confident that the true mean purchase amount across all transactions falls between Rs 9,250.69 and Rs 9,277.24.



Sample Size:	550,068
Mean Purchase:	Rs 9,263.97
Std Deviation:	Rs 5,023.07
Standard Error:	Rs 6.7727
95% CI Lower:	Rs 9,250.69
95% CI Upper:	Rs 9,277.24

Interpretation: We are 95% confident the true mean purchase amount falls between Rs 9,251 and Rs 9,277.

- **Hypothesis Testing:** A significant difference in spending was found between genders. Males (mean: Rs 9,437.53) spent significantly more than females (mean: Rs 8,734.57). No significant difference was found between married and unmarried customers.

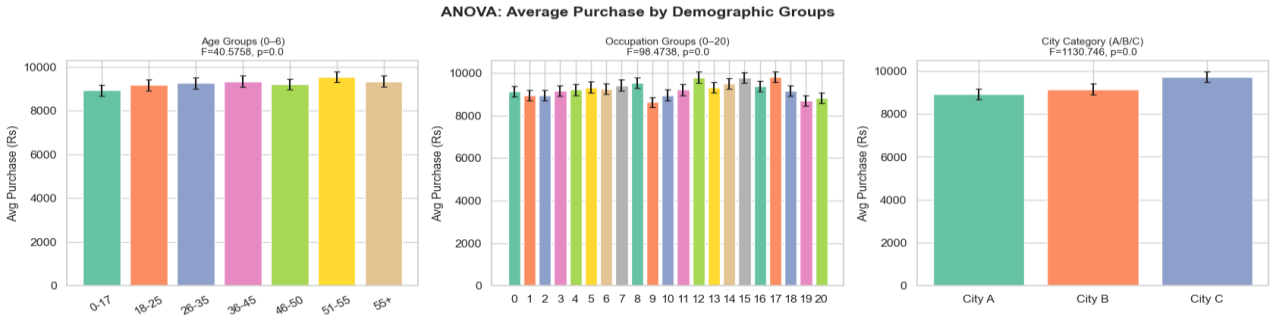
IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

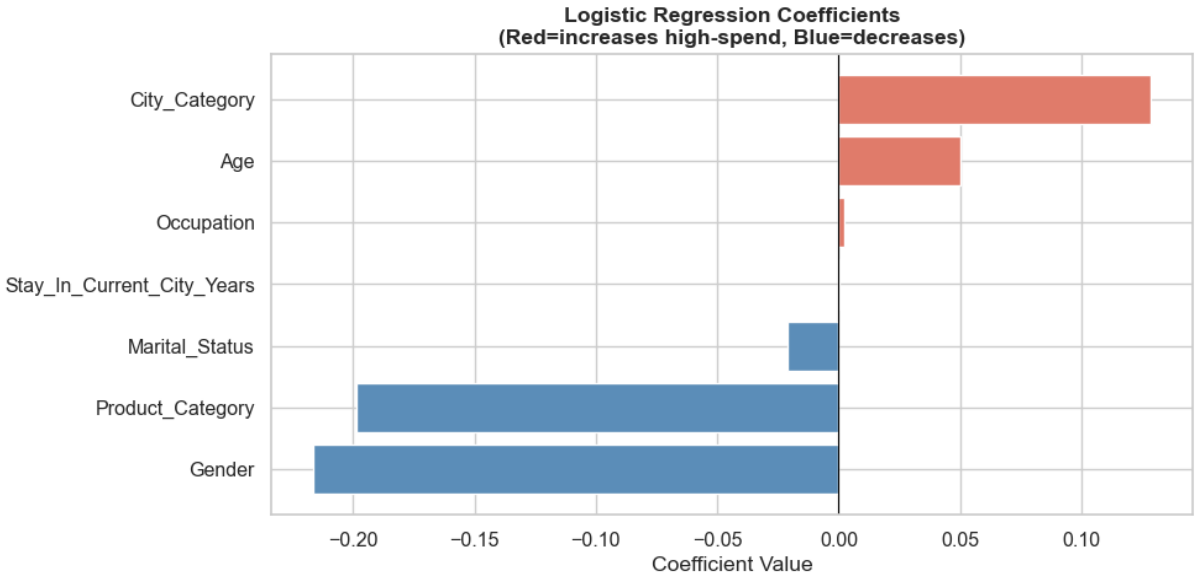
Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14410

- **ANOVA:** All demographic variables—Age, Occupation, City Category, Stay in Current City Years, and Product Category—showed a statistically significant ($p < 0.05$) impact on the average purchase amount. City Category and Product Category had the most substantial influence.



- **Logistic Regression:** This model achieved a test accuracy of 78.87%. The coefficients revealed that being male, older, and having a higher occupation code were among the factors that increased the probability of being a high spender (purchase amount ≥ Rs 9,263).



2. Machine Learning Models (Predictive Lens)

K-Means Clustering: This unsupervised technique segmented customers into three distinct groups: High, Medium, and

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14410

Low Spenders, based on their demographic attributes and purchase history. The average purchase values for these clusters were Rs 9,453.76, Rs 9,413.82, and Rs 8,734.57, respectively.

Model	Accuracy	F1 Score
XGBoost	86.60%	0.8664
LightGBM	86.54%	0.8660
Gradient Boosting	86.53%	0.8658
Decision Tree	86.47%	0.8653
Random Forest	84.81%	0.8479

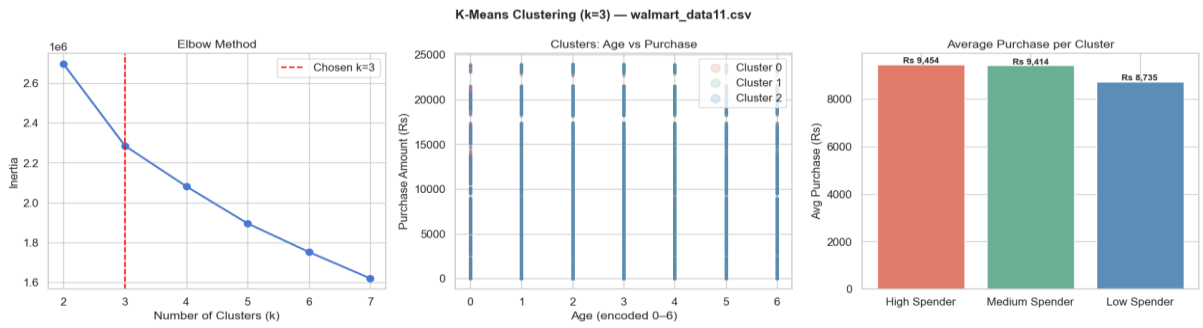
Table [1]: Accuracy

- Classification Models: The goal was to predict whether a customer would be a "High Spender" (≥ Rs 9,263). The results are as follows:
- Regression Models: The goal was to predict the exact purchase amount. The results are as follows:

Model	RMSE	MAE	R^2
Decision Tree	5025.74	4066.78	-0.0022
LightGBM	5026.30	4066.23	-0.0024
Gradient Boosting	5027.95	4066.73	-0.0031
Random Forest	5028.90	4067.13	-0.0035
XGBoost	5043.77	4074.73	-0.0094

Table [2]: Error Values

Key Insight on Regression: All models performed poorly for regression, with R^2 values near zero or negative. This suggests that the demographic features alone are insufficient to predict the exact purchase amount with any meaningful accuracy, highlighting the need for additional features like product details and timing of purchase for this specific task**.**



IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

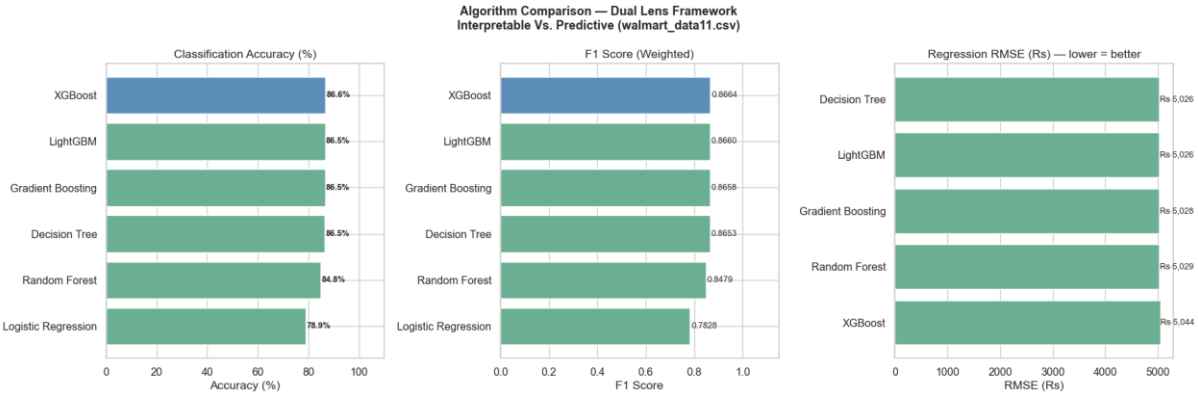
DOI: 10.17148/IJIREEICE.2026.14410

3. Dual-Lens Synthesis

- Interpretability: The statistical analysis clearly identified *which* demographic factors (Gender, Age, Occupation, City, Marital Status) influence spending, providing a transparent explanation of "why" certain groups spend more. Logistic regression offered a transparent, rule-based understanding of these influences.

- Predictive Power: The machine learning models, particularly XGBoost, demonstrated a significant improvement in predictive accuracy (86.6% vs. 78.9%) for the classification task of identifying high-value customers.

- Complementarity: The combined approach reveals that while complex ML models are superior for prediction, the insights from statistical methods are crucial for understanding the underlying drivers and forming actionable business strategies. For example, while we know *why* older males are high spenders, XGBoost can more accurately identify *which* older males will spend the most.



V.CONCLUSION

This research successfully implemented a dual-lens framework to analyze Walmart's Black Friday consumer behavior, demonstrating that statistical and machine learning methods are not mutually exclusive but are, in fact, complementary tools. The statistical analysis provided a transparent and interpretable understanding of the demographic drivers of spending, confirming that gender, age, and city category are significant factors. The machine learning models, specifically XGBoost, proved superior for the predictive task of identifying high-spending customers, achieving 86.6% accuracy.

The findings underscore a critical insight: for this dataset, demographic features alone are sufficient for effective classification (predicting high vs. low spenders) but are inadequate for accurate regression (predicting the exact purchase amount). The negative R^2 values for regression models highlight the complexity of the target variable, suggesting that factors beyond simple demographics-such as product attributes, emotional state, or in-the-moment decision-making-are key drivers of the final purchase value.

By integrating these two analytical lenses, this study provides a robust framework that offers both the "why" and the "who" for retail analytics. This integrated approach empowers businesses to make data-driven decisions that are both strategically sound (based on causal understanding) and operationally precise (based on predictive accuracy). The results offer a clear roadmap for Walmart to refine its marketing, pricing, and inventory strategies for future Black Friday events.

VI.FUTURE ENHANCEMENT

While this research provides a strong foundation, several avenues for future work could yield even deeper insights and more robust models.

1. Incorporate Product and Temporal Data: The most impactful enhancement would be to include more detailed product information (e.g., product price, brand, category hierarchy) and time-based features (e.g., hour of purchase, day of the week). This data is crucial for improving regression models, which performed poorly with only demographic data, and could reveal time-sensitive spending patterns.

DOI: 10.17148/IJIREEICE.2026.14410

2. Deep Learning and Advanced Neural Networks: Exploring deep learning architectures like Deep Neural Networks (DNNs) or TabNet could help capture complex, non-linear relationships that tree-based models might miss. Recurrent Neural Networks (RNNs) or LSTMs could be used if the data were structured as a time series of individual customer purchases over the Black Friday period.

3. Real-Time Analytics Platform: The current analysis is static. A future enhancement would be to develop a scalable data pipeline and real-time analytics platform. This would allow for live predictions and insights during the Black Friday event, enabling dynamic pricing, real-time inventory reallocation, and instantaneous personalized offers to customers based on their behavior as it unfolds.

4. Advanced Customer Segmentation: While K-Means was used, more advanced clustering techniques (e.g., Hierarchical Clustering, DBSCAN, Gaussian Mixture Models) could be explored to identify more nuanced and natural customer segments. Combining clustering with supervised learning in a two-stage model could further refine predictions.

5. Explainable AI (XAI) for ML Models: To bridge the interpretability gap further, applying XAI techniques like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) to the best-performing ML models (e.g., XGBoost) would provide model-specific, per-prediction explanations. This would allow us to understand *why* a specific model flagged a particular customer as a "high spender," combining the best of both worlds.

6. Cross-Event Analysis: Extend the analysis to compare consumer behavior across different sales events (e.g., Black Friday vs. Cyber Monday vs. regular weekly sales) to understand how discount-driven urgency affects spending patterns.

REFERENCES

[1]. Kelly, Scott & Bowman (2025) - Show how promotions influence consumer behavior, linking price elasticity and marginal utility to spending.
[2]. Bakhrun et al. (2025) - Use visualization to highlight Walmart consumer trends, reinforcing data preparation.
[3]. Neba et al. (2024) - Demonstrate predictive strength of Random Forest, GBM, LightGBM, and XGBoost in retail forecasting.
[4]. Patil et al. (2023) - Highlight XGBoost regression effectiveness with careful data preparation and fair evaluation.
[5]. Vyavahare et al. (2023) - Show how demographic features improve prediction accuracy with Random Forest.
[6]. Ali & Shah (2024) - Emphasize hybrid ML + time series approaches, proving the value of diverse data.
[7]. Elsbree (2022) - Analyze Walmart's Black Friday pricing behavior, supporting interpretability through statistical insights.
[8]. Dr. Akhilesh, A. Waoo (2023) - Customer Behavior Analysis in E-Commerce using Machine Learning Approach: A Survey.
[9]. Ridwan, Ishola Bayo (2025) - Transforming Customer Segmentation with Unsupervised Learning Models and Behavioral Data in Digital Commerce.

