Journal of Artificial Intelligence in Governance and Public Policy (JAIGPP)
Volume 1 | Issue 2 | April-June 2026 | Pages 23–29
DOI: https://doi.org/xxxxxxxxxxxxx

Research Article OPEN ACCESS Available online: https://jaigpp.org/index.php/j

Machine Learning-Based Customer Churn Prediction in E-Commerce Platforms Using Behavioral Data Analytics

Dr. Pradeep Laxkar
Associate Professor, Department of Computer Science and Engineering, ITM (SLS) University, Vadodara, Gujarat
pradeep.laxkar@gmail.com
Corresponding Author: *Pradeep Laxkar (pradeep.laxkar@gmail.com)

Received: 15 April 2026 | **Revised:** 20 may 2026 | **Accepted:** 15 June 2026 |

***Abstract*—Customer churn prediction has become a critical challenge for e-commerce organizations as retaining existing customers is more cost-effective than acquiring new ones. This study proposes a machine learning-based customer churn prediction framework using behavioral data analytics to identify customers at risk of leaving an e-commerce platform. The proposed approach uses the Kaggle E-Commerce Customer Churn dataset and incorporates comprehensive data preprocessing, including missing-value handling, categorical feature encoding, feature scaling, and class balancing via the Synthetic Minority Over-sampling Technique (SMOTE). Recursive Feature Elimination (RFE) is employed to select the most relevant features, while K-Means clustering is used for customer segmentation. A Light Gradient Boosting Machine (LightGBM) classifier is then developed to predict customer churn, and Shapley Additive explanations (SHAP) are applied to improve model interpretability by explaining feature contributions. Experimental results demonstrate that the proposed model achieves 96.00% accuracy, 95.00% precision, 92.00% recall, 93.00% F1-score, and a ROC-AUC score of 95.1%, outperforming several existing approaches. The proposed framework provides an accurate, efficient, and explainable solution to support proactive customer retention strategies and improve long-term business profitability.**

Keywords*— *Customer Churn Prediction, LightGBM, Behavioral Data Analytics, Explainable Artificial Intelligence (SHAP), E-Commerce.

I. INTRODUCTION

The rapid growth of e-commerce and digital commerce has intensified market competition, making customer retention a strategic priority for businesses seeking sustainable growth, profitability, and long-term customer relationships [1][2]. Customer churn refers to the phenomenon in which customers discontinue purchasing products or services from a business and switch to competing alternatives. In the highly competitive online shopping industry [3], customer retention has become as important as customer acquisition because retaining existing customers is significantly more cost-effective than attracting new ones [4]. High customer churn negatively impacts business profitability, sales, customer lifetime value, and market competitiveness [5][6][7]. Therefore, accurately identifying customers who are likely to leave an e-commerce platform has become a critical objective for organizations seeking sustainable growth and improved customer relationship management [8][9].

The rapid growth of digital commerce has resulted in the generation of massive volumes of customer behavior data, including purchase history, browsing activities, transaction frequency, product preferences, session duration, cart abandonment patterns, and customer engagement records [10][11]. These behavioral data provide valuable insights into customers' purchasing intentions and loyalty. Behavioral data analytics enables businesses to transform these raw data into meaningful patterns that reveal changes in customer behavior, allowing organizations to detect early signs of churn and develop proactive customer retention strategies.

Recent advances in machine learning have significantly improved the ability to predict customer churn by learning complex relationships from large-scale customer datasets. Unlike traditional statistical approaches, machine learning algorithms can effectively model nonlinear interactions among multiple behavioral attributes while maintaining high predictive performance on high-dimensional data. Supervised learning techniques have therefore become widely adopted for churn prediction because of their capability to identify subtle behavioral patterns that distinguish churn-prone customers from loyal ones.

In e-commerce platforms, accurate churn prediction supports personalized marketing campaigns, targeted promotional offers, and customer engagement strategies that enhance customer satisfaction and long-term loyalty [12]. By leveraging behavioral data analytics together with advanced machine learning techniques, organizations can make data-driven decisions that reduce customer attrition and maximize business revenue [13]. Therefore, this study presents a machine learning-based customer churn prediction framework for e-commerce platforms that utilizes customer behavioral data, incorporates feature selection and customer segmentation, and employs the LightGBM classifier to accurately identify customers at risk of churn, thereby supporting effective retention strategies and improving overall business performance.

Journal of Artificial Intelligence in Governance and Public Policy, 2(6):23-29
DOI: https://doi.org/xxxxxxxxxxxxx

A. Motivation and Contributions of the Study

The motivation for this research arises from the increasing competition among e-commerce platforms and the growing challenge of customer retention. Acquiring new customers is significantly more expensive than retaining existing ones, making customer churn prediction a critical business objective. Traditional statistical approaches often fail to capture complex behavioral patterns hidden within large-scale customer data, resulting in lower prediction accuracy. With the rapid growth of machine learning techniques, it has become possible to analyse customer behavioural, transactional, and demographic information more effectively, enabling businesses to identify customers at risk of churn and implement timely retention strategies. The major contributions of this study are as follows:

- Utilized the Kaggle E-Commerce Customer Churn Dataset containing customer demographic, transactional, and behavioral information.
- Performed comprehensive data preprocessing, including missing value handling, categorical feature encoding, feature scaling, and class balancing using SMOTE.
- Applied Recursive Feature Elimination (RFE) for selecting the most informative customer behavioral features.
- Performed customer segmentation using K-Means clustering to group customers with similar purchasing behavior.
- Developed a LightGBM-based customer churn prediction model and enhanced model interpretability using SHAP (SHapley Additive exPlanations).
- Evaluated the proposed framework using Accuracy, Precision, Recall, F1-score, ROC-AUC, Confusion Matrix, and Feature Importance Analysis, demonstrating its effectiveness for customer churn prediction.

B. Justification and Novelty of the Paper

This study justifies the application of advanced machine learning techniques for customer churn prediction in e-commerce platforms. Conventional retention methods often fail to capture complex customer behavioral patterns. The proposed framework integrates LightGBM, SMOTE, Recursive Feature Elimination (RFE), K-Means clustering, and SHAP to develop an accurate and explainable churn prediction model. Its novelty lies in combining feature selection, customer segmentation, explainable artificial intelligence, and gradient boosting within a unified framework, enabling accurate prediction and supporting more effective customer retention strategies.

Organization of the paper

The paper is organized as follows: Section II presents the literature review on customer churn prediction in e-commerce. Section III describes the proposed methodology. Section IV discusses the experimental results. Finally, Section V concludes the paper and outlines future research directions.

II. Literature Review

The literature review highlights recent machine learning approaches for customer churn prediction, emphasizing predictive performance, feature engineering, customer segmentation, and recommendation techniques.

I. Jahan and T. F. Sanam, (2026), CatBoost achieves the best customer churn prediction performance based on accuracy and F1-score. Recursive Feature Elimination ranks features for customer segmentation. K-means outperforms Hierarchical Clustering and DBSCAN, supported by a Hopkins score of 0.0932. Collaborative filtering performs better than popularity-based recommendations, while the SVD model achieves an RMSE of 1.26, MAE of 0.99, MSE of 1.69, and 89% precision with the shortest test time [14].

H. Ren (2025), utilizes the Kaggle E-commerce Customer Churn dataset. Missing values are imputed using KNN, followed by feature engineering and preprocessing. Multiple models, including Logistic Regression, Random Forest, GBDT, Neural Network, and a stacking model with XGBoost as the meta-learner, are developed. The stacking model achieves the best performance with 92.8% accuracy and 0.940 AUC [15].

A. Budiyono and I. Nendi, (2025), A machine learning-based predictive model identifies high-risk Shopee customers in Indonesia using behavioural and transactional data. The dataset includes transaction frequency, recency, voucher usage, session count, and promotional interactions. SMOTE addresses data imbalance, and XGBoost achieves the best performance with an AUC of 0.948, demonstrating strong discriminative ability [16].

M. S. Hasan et al., (2024), develops machine learning models to identify customers likely to churn using behavioral patterns, transaction history, and demographic data. A systematic training and testing strategy is adopted. XGBoost achieves the best overall performance, with all evaluation metrics above 0.90, while Random Forest ranks second, with scores generally above 0.85 [17].

A. Khattak et al., (2023), provided an effective method for predicting customer churn based on a hybrid deep learning model termed BiLSTM-CNN. The goal is to effectively estimate customer churn using benchmark data and increase the churn prediction process's accuracy. The experimental results show that when trained, tested, and validated on the benchmark dataset, the proposed BiLSTM-CNN model attained a remarkable accuracy of 81% [18].

L. Gan, (2022), the improved XGBoost algorithm can effectively reduce the probability of class I errors and has higher accuracy, among which the accuracy has increased by 2.8%. The prediction effect of customer groups after segmentation was better than that before segmentation, in which the probability of the occurrence of class I errors in the prediction of core value customers decreases by 10.8% and the accuracy rate increases by 7.8% [19].

Table I summarizes the reviewed studies, comparing their approaches, datasets, key findings, limitations, and future research recommendations to identify existing research gaps.

DOI: https://doi.org/xxxxxxxxxxxxx

TABLE I. COMPARATIVE REVIEW OF MACHINE LEARNING TECHNIQUES FOR CUSTOMER CHURN PREDICTION

Author	Approach	Dataset	Findings	Limitations	Recommendations
I. Jahan and T. F. Sanam (2026)	CatBoost, RFE, K-means, Collaborative Filtering, SVD	E-commerce customer behavioral dataset	CatBoost achieved the best churn prediction; K-means and SVD showed superior performance.	Limited evaluation on diverse datasets and real-time deployment.	Validate the framework on large-scale and real-world e-commerce platforms.
H. Ren (2025)	KNN, Feature Engineering, LR, RF, GBDT, NN, Stacking (XGBoost)	Kaggle E-commerce Customer Churn Dataset	Stacking model achieved 92.8% accuracy and 0.940 AUC.	Focused only on supervised learning without customer segmentation.	Integrate customer segmentation and recommendation mechanisms.
A. Budiyono and I. Nendi (2025)	SMOTE and XGBoost	Shopee customer behavioral dataset	XGBoost achieved 0.948 AUC for churn prediction.	Limited to a single e-commerce platform.	Evaluate model generalization across multiple e-commerce platforms.
M. S. Hasan et al. (2024)	XGBoost and Random Forest	Behavioral, transactional, and demographic data	XGBoost outperformed Random Forest with metrics above 0.90.	No feature selection or customer clustering analysis.	Incorporate feature optimization and customer segmentation.
A. Khattak et al. (2023)	BiLSTM-CNN Hybrid Deep Learning	Benchmark churn dataset	Achieved 81% accuracy in churn prediction.	Moderate accuracy and high computational complexity.	Explore lightweight hybrid and ensemble learning models.
L. Gan (2022)	Improved XGBoost with Customer Segmentation	Customer segmentation dataset	Improved XGBoost increased accuracy by 2.8% after segmentation.	Limited comparison with recent ensemble techniques.	Compare with advanced boosting and deep learning models.

III. PROPOSED METHODOLOGY

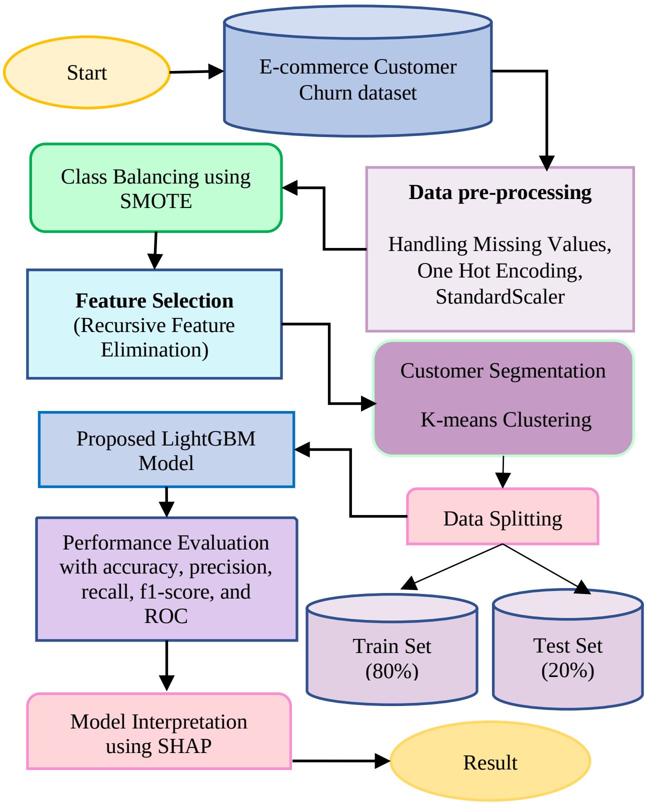

Fig. 1. Proposed Methodology for Customer Churn Prediction in E-Commerce Platform

The proposed methodology utilizes the E-commerce Customer Churn dataset, where data preprocessing is performed using missing value handling, one-hot encoding, and StandardScaler. SMOTE is applied to balance the dataset, followed by Recursive Feature Elimination (RFE) for feature selection and K-means clustering for customer segmentation. The dataset is then split into 80% training and 20% testing sets, and a LightGBM model is trained for churn prediction. Finally, the model is evaluated using the confusion matrix, accuracy, precision, recall, F1-score, ROC-AUC, and SHAP analysis. The proposed workflow is shown in Figure 1.

The whole steps of this methodology are described below

A. Data Collection

The proposed study utilizes the E-Commerce Customer Churn Dataset obtained from Kaggle [20]. The dataset contains customer demographic information, transaction history, purchasing behavior, application usage statistics, satisfaction scores, cashback information, complaints, payment methods, and other behavioral attributes. The target variable is Churn, where customers are categorized as retained (0) or churned (1). These behavioral features provide valuable information for developing an accurate churn prediction model.

B. Data Preprocessing

The E-Commerce dataset used for this particular customer churn classification consisted of 5630 samples with 20 attributes. The presence of null values in the dataset would have made it difficult for us to process the data, so they were dealt by adding median values of the data frame of other columns to avoid null values in the dataset. Also, various categorical variables that were set as numerical variables were converted to process the data. Categorical variables are transformed into numerical representations through Label Encoding and One-Hot Encoding wherever appropriate.

Unlike conventional preprocessing approaches that employ Min-Max normalization, this study applies StandardScaler to standardize numerical features by transforming them into zero mean and unit variance. Standardizing the data ensures that each feature contributes equally to the model. The Z-score standardization process involves transforming each feature x′ using the formula (1):

$$x' = \frac{x - \mu}{\sigma} \quad (1)$$

C. Handling Class Imbalance

Imbalanced data are a common challenge in data mining, where the majority class significantly outnumbers the

DOI: https://doi.org/xxxxxxxxxxxxx

minority class. The Synthetic Minority Over-Sampling Technique (SMOTE) balances the dataset by generating synthetic instances of the minority class [21]. This process improves the classifier's ability to correctly identify churned customers while reducing prediction bias.

D. Feature Selection using Recursive Feature Elimination (RFE)

Feature selection is to select a subset of relevant features from a larger set of original ones in terms of some pre-defined criteria such as classification performance or class separability [22]. To reduce data dimensionality and improve prediction performance, Recursive Feature Elimination (RFE) is employed for feature selection. RFE recursively removes less important features based on model importance until the optimal subset of behavioral variables is obtained. Selecting the most informative features improves computational efficiency and enhances the predictive capability of the proposed churn prediction model.

E. Customer Segmentation using K-Means Clustering

One of the tasks of customer churn management is to divide customer groups into various types according to their shopping behaviors, such as valuable core customers, which are loyal customers and easy-to-lose customers. To formulate targeted marketing strategies for customers, enterprises must subdivide customers [23]. In this study, k-means was used for customer segmentation. As a classical algorithm of unsupervised clustering, k-means is similar to a fully automatic classification.

F. Data Splitting

Data splitting divides the dataset into training and testing sets to evaluate model performance and prevent overfitting. In this study, 80% of the data is used for training and 20% for testing.

G. LightGBM Model Development

The proposed customer churn prediction model is developed using the Light Gradient Boosting Machine (LightGBM) algorithm. LightGBM is selected because of its high computational efficiency, fast training speed, ability to handle large-scale datasets, and strong predictive performance for classification problems. After completing data preprocessing, feature selection, and customer segmentation, the dataset is divided into training and testing sets using an 80:20 ratio. The LightGBM classifier is trained using the training dataset and evaluated on the testing dataset to classify customers as churned or retained.

H. Model Evaluation

In the customer churn prediction task, the evaluation of the model performance can be achieved using a confusion matrix, which consists of four key components: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

1) Accuracy

Accuracy refers to the proportion of samples correctly predicted by the model to the total samples, as shown in Equation 2. For the churn prediction task, Accuracy is used to measure the overall prediction accuracy of the model.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

2) Precision

This metric measures the proportion of correctly predicted churned customers out of all customers predicted to churn [24]. High precision indicates that the model is effective at minimizing false positives, which is crucial for targeting retention efforts. It is derived in Equation (3):

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

3) Recall

Recall is the proportion of actual churn customers successfully predicted as churn customers by the model, as shown in Equation (4). In churn prediction, a higher Recall indicates that the model performs well in capturing real churn customers and has a lower underreporting rate.

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

4) F1-Score

F1 is the reconciled average of Precision and Recall, as shown in Equation (5). F1 integrates the model’s prediction accuracy and ability to capture real churn customers, and can effectively assess the overall performance of the model.

$$F1 = \frac{2\,TP}{2\,TP+FP+FN} \tag{5}$$

5) AUC-ROC Score

The Area Under the Receiver Operating Characteristic curve (AUC-ROC) evaluates the model's ability to distinguish between churned and non-churned customers. Mathematically, the AUC is defined as the area under the ROC curve and is calculated as (6):

$$\mathrm{AUC-ROC} = \int_0^1 \mathrm{TPR}\,(\mathrm{t})\mathrm{d}t \tag{6}$$

A higher AUC value indicates better performance in classifying customers correctly, enhancing the model's overall effectiveness in churn prediction.

I. Model Interpretation using SHAP

After training the LightGBM classifier, SHapley Additive exPlanations (SHAP) are employed to interpret the model predictions. SHAP quantifies the contribution of each behavioral feature toward customer churn prediction at both the global and individual levels. SHAP Summary Plots, Feature Importance Bar Plots, and Dependence Plots are generated to identify the most influential behavioral features affecting churn. The integration of SHAP enhances the transparency and interpretability of the LightGBM model, enabling businesses to better understand the factors influencing customer churn and supporting informed decision-making.

IV. RESULTS AND DISCUSSION

The proposed customer churn prediction framework was evaluated using an Intel Core i7 processor with 64 GB RAM, Windows 10, Jupyter Notebook, and Python. The framework integrates data preprocessing, SMOTE for class balancing, Recursive Feature Elimination (RFE), K-Means clustering, a LightGBM classifier, and SHAP for model interpretability. Experimental results demonstrate that the proposed framework effectively distinguishes churned and retained customers with high prediction accuracy and low classification error. The following sections present a detailed performance analysis using evaluation metrics, confusion

 Journal of Artificial Intelligence in Governance and Public Policy, 2(6):23-29
DOI: https://doi.org/xxxxxxxxxxxxx

matrix, ROC curve, SHAP-based feature interpretation, and comparative analysis.

TABLE II. PERFORMANCE EVALUATION OF LIGHTGBM CLASSIFIER

Performance Metric	Value
Accuracy	**96.00**
Precision	**95.00**
Recall	**92.00**
F1-Score	**93.00**
ROC-AUC Score	**95.1**

In Table II, the performance of the proposed LightGBM model is presented in customer churn prediction. The model attained an accuracy of 96%, precision of 95%, recall of 92%, F1-score of 93%, and a ROC-AUC score of 95.1%, showing that the model performs highly and is capable of identifying churning customers with high accuracy while balancing precision and recall well.

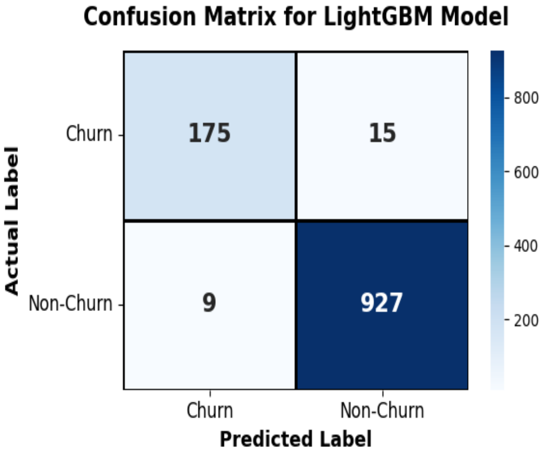

Fig. 2. Confusion Matrix of the LightGBM Classifier

The classifier correctly identified 175 customers who churned and 927 who were retained. It only misclassified 15 churn customers as non-churn and 9 retained customers as churn as shown in Figure 2. The low number of false positives and false negatives indicates that the proposed model is effective in identifying customers at risk of churn from loyal customers. In conclusion, the confusion matrix reveals that the proposed LightGBM classifier offers robust and stable prediction results with minimal classification errors.

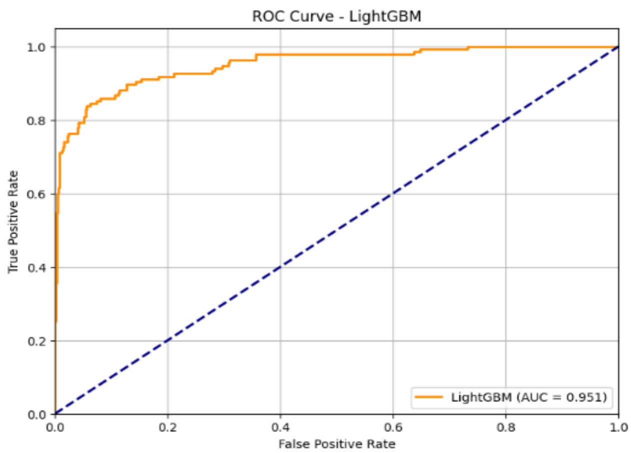

Fig. 3. ROC Curve of the LightGBM Classifier

Figure. 3 presents the ROC curve of the proposed LightGBM model for customer churn prediction. The model achieved a ROC-AUC score of 0.951, indicating excellent discriminative ability between churn and non-churn customers. The ROC curve remains close to the upper-left corner, demonstrating a high true positive rate with a low false positive rate, which reflects the model's strong classification performance.

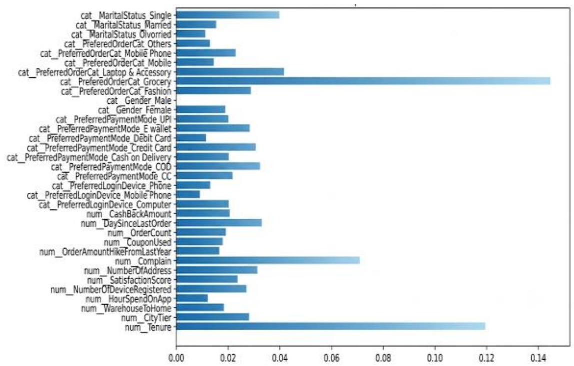

Fig. 4. Feature Importance of the Proposed LightGBM Classifier

Figure 4 is the representation of SHAP feature importance for the developed LightGBM model. The findings reveal that PreferredOrderCat_Grocery, Tenure, and Complain are among the important features used to predict customer churn. The differences in importance indicate that customer behavior, loyalty, and complaint behavior are important factors in predicting customer churn.

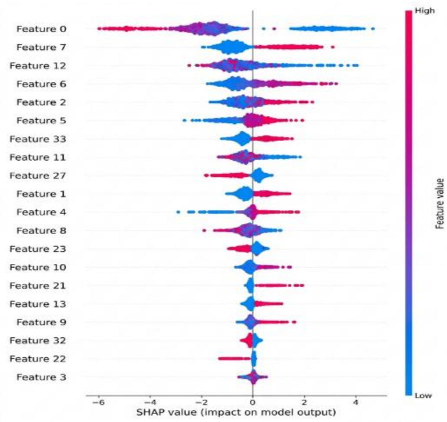

Fig. 5. SHAP Value

SHAP summary plot for the LightGBM classifier is given in Figure 5. The SHAP summary plot shows the contribution of each feature toward the predicted output. This plot represents the ranking of features according to their impact on churn prediction, in which the top features have maximum influence on the output. The plot shows the contribution of low and high values of features in prediction.

A. *Comparative Analysis*

A comparative analysis was conducted between the proposed LightGBM classifier and several existing machine learning approaches for customer churn prediction using different benchmark datasets, as presented in Table III. The proposed LightGBM method demonstrated a classification accuracy of 96.00%, with precision of 95.00%, recall of 92.00%, F1-score of 93.00%, and an excellent ROC-AUC of 95.1%.

The proposed model outperforms all the previous models such as XGBoost, SVM, logistic regression, and K-Nearest Neighbours (KNN) based on all the metrics. In this case, the model outperforms all the traditional models on the basis of precision and F1 score, whereas its discriminative power is also very high based on ROC-AUC score.

TABLE III. COMPARATIVE EVALUATION OF CUSTOMER CHURN PREDICTION MODELS

Reference	Dataset	Model	Accuracy	Precision	Recall	F1-Score	ROC AUC
Proposed Method	E-commerce churn dataset	LightGBM	96.00	95.00	92.00	93.00	95.1
[25]	Telecom Customer Churn Dataset	XG Boost Classifier	96.2	54.00	83.00	83.6	79.00
[26]	Ecommerce Customer Churn Analysis and Prediction	SVM	89.87	76.27	51.15	51.15	74.08
[27]	Olist Online dataset	Logistic Regression	92.11	93.49	90.80	92.12	89.72
[28]	E-commerce churn dataset	K-Neighbors Classifier	92.36	87.14	64.21	73.94	94.66

Better performance of the proposed framework is due to the use of data preprocessing, class balancing through SMOTE, feature selection using RFE, customer segmentation with K-means clustering and lightGBM's efficient learning ability. In addition to this, inclusion of SHAP helps improve the interpretability of the model by understanding the contribution of the behavioral features of the customer.

Overall, the experimental results proves that the proposed LightGBM-based customer churn prediction framework provides an accurate, computationally efficient, and explainable solution for identifying customers at risk of churn. The framework allows e-commerce organizations to implement proactive customer retention strategies, improve customer satisfaction, and maximize long-term business profitability.

V. CONCLUSION AND FUTURE SCOPE

Customer churn prediction has become an essential component of customer relationship management in modern e-commerce platforms. This study proposed a machine learning-based customer churn prediction framework using customer behavioral data analytics. The framework integrates data preprocessing, SMOTE for class balancing, Recursive Feature Elimination (RFE) for feature selection, K-Means clustering for customer segmentation, the LightGBM classifier for churn prediction, and SHAP for model interpretability. Experimental results demonstrated that the proposed framework achieved 96.00% accuracy, 95.00% precision, 92.00% recall, 93.00% F1-score, and a 95.1% ROC-AUC score, indicating excellent predictive performance. Furthermore, SHAP analysis identified the most influential behavioral factors contributing to customer churn, improving model transparency and interpretability. Overall, the proposed framework provides an accurate, computationally efficient, and explainable solution for identifying customers at risk of churn, thereby enabling e-commerce organizations to implement proactive retention strategies, enhance customer satisfaction, and improve long-term business profitability.

Future research may investigate deep learning and transformer-based approaches, real-time customer churn prediction using streaming behavioral data, multi-platform e-commerce datasets, automated hyperparameter optimization, and personalized retention recommendation systems to further improve prediction accuracy, scalability, and practical deployment.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The author declares no conflict of interest.

P. Laxkar, "Machine Learning-Based Customer Churn Prediction in E-Commerce Platforms Using Behavioral Data Analytics," *Journal of Artificial Intelligence in Governance and Public Policy (JAIGPP)*, vol. 1, no. 2, pp. 23–29, Apr.–Jun. 2026, doi: xx.xxxx/xxxxx.

REFERENCES

[1] A. Joon, S. A. Pahune, S. Mathur, and S. Rongala, "Designing Predictive Analytics Frameworks with ML for ERP-Supported E-Commerce Systems," in *2025 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*, 2025, pp. 1–6. doi: 10.1109/ASSIC64892.2025.11158452.

[2] T. Shah, "IJRAR published paper measuring profitibility BNPL customer analytics," *Int. J. Res. Anal. Rev.*, vol. 10, no. 1, pp. 505–510, February, 2023, doi: 10.56975/ijrar.v10i1.322189.

[3] N. Bhatt, "A Hybrid Artificial Intelligence Model for Product Recommendation in Online Retail System," in *Artificial Intelligence Systems (AIS 2026)*, 2026, pp. 1–7.

[4] Y. Patel, "A Context-Aware Recommendation Model for Improving Conversion Rates in E-Commerce Platforms," *Am. J. Data Sci. Artif. Intell. Innov.*, vol. 5, pp. 58–80, Januray, 2025.

[5] Shanthini B, "Predicting Customer Churn Using Artificial Neural Networks: A Data-Driven Approach to Enhance Customer Retention in E-Commerce," *J. Inf. Syst. Eng. Manag.*, vol. 10, no. 41s, pp. 380–388, Apr. 2025, doi: 10.52783/jisem.v10i41s.7854.

[6] H. Ravilla, "Building Scalable Applications with Heroku and Salesforce Integration," *Am. J. Technol.*, vol. 4, no. 3, pp. 15–36, Dec. 2025, doi: 10.58425/ajt.v4i3.454.

[7] K. Dixit, "Predictive Analytics in Business Intelligence for Sales Forecasting," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 3, no. 1, pp. 1–11, 2023, doi: 10.48175/IJARSCT-12750G.

[8] S. Kukar, "Explainable AI Models for Transparent and Ethical Customer Relationship Management Decision," in *2026 IEEE 2nd International Conference on Secure IoT, Assured and Trusted Computing (SATC)*, Houston, TX, USA: IEEE, 2026, pp. 1–7, March. doi: 10.1109/SATC69565.2026.11542554.

[9] J. Kofi, "Data-Driven Decision Support for CRM Campaigns in the Banking Sector Through Supervised Learning Models," in *2026 International Conference on Communication, Computing and Emerging Technologies (IC3ET)*, Vasai, India: IEEE, 2026, pp. 1–8, February. doi: 10.1109/IC3ET64989.2026.11467470.

[10] G. Mena, K. Coussement, K. W. De Bock, A. De Caigny, and S. Lessmann, "Exploiting time-varying RFM measures for customer churn prediction with deep neural networks," *Ann. Oper. Res.*, vol. 339, no. 1–2, pp. 765–787, Aug. 2024, doi: 10.1007/s10479-023-05259-9.

[11] Harpreet Singh, "The Metaverse Revolution: Transforming the Future of Retail," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 11, no. 2, pp. 109–115, Mar. 2025, doi: 10.32628/CSEIT251112391.

[12] Y. Patel, "A Trust-Aware Authentication Framework for Preventing Account Takeover Attacks in E-Commerce Platforms," *Newark J. Human-Centric AI Robot. Interact.*, vol. 5, pp. 287–320, jan, 2025.

[13] L. Li, "Research on Improved Xgboost Algorithm for Big Data Analysis of e-Commerce Customer Churn," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 12, pp. 1086–1094, 2022, doi: 10.14569/IJACSA.2022.01312124.

[14] I. Jahan and T. F. Sanam, "A comprehensive framework for customer retention in E-commerce using machine learning based on churn prediction, customer segmentation, and recommendation," *Electron. Commer. Res.*, vol. 26, no. 1, pp. 1–44, 2026, doi: 10.1007/s10660-024-09936-0.

[15] H. Ren, "Machine Learning-Based Prediction of Customer Churn Risk in E-commerce," 2025, pp. 47–52. doi: 10.54254/2754-1169/153/2024.19473.

[16] A. Budiyono and I. Nendi, "Customer Churn Prediction Uses Machine Learning to Improve Retention on Digital Platforms," *J. Digit. Bus. Data Sci.*, vol. 2, no. 2, pp. 55–75, Dec. 2025, doi: 10.59261/jdbs.v2i2.23.

[17] M. S. Hasan *et al.*, "Predictive Analytics for Customer Retention: Machine Learning Models to Analyze and Mitigate Churn in E-Commerce Platforms," *J. Bus. Manag. Stud.*, vol. 6, no. 4, pp. 304–320, 2024.

[18] A. Khattak, Z. Mehak, H. Ahmad, M. U. Asghar, M. Z. Asghar, and A. Khan, "Customer churn prediction using composite deep learning technique," *Sci. Rep.*, 2023, doi: 10.1038/s41598-023-44396-w.

[19] L. Gan, "XGBoost-Based E-Commerce Customer Loss Prediction," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–10, Jul. 2022, doi: 10.1155/2022/1858300.

[20] Ankit Verma, "Ecommerce Customer Churn Analysis and Prediction," Kaggle.

[21] M. Imani and H. R. Arabnia, "Hyperparameter Optimization and Combined Data Sampling Techniques in Machine Learning for Customer Churn Prediction: A Comparative Analysis," *Technologies*, vol. 11, no. 6, p. 167, Nov. 2023, doi: 10.3390/technologies11060167.

[22] X. Chen and J. C. Jeong, "Enhanced recursive feature elimination," in *Sixth International Conference on Machine Learning and Applications (ICMLA 2007)*, IEEE, Dec. 2007, pp. 429–435. doi: 10.1109/ICMLA.2007.35.

[23] X. Xiahou and Y. Harada, "B2C E-Commerce Customer Churn Prediction Based on K-Means and SVM," *J. Theor. Appl. Electron. Commer. Res.*, vol. 17, no. 2, pp. 458–475, Apr. 2022, doi: 10.3390/jtaer17020024.

[24] M. A. Shaikhsurab and P. Magadum, "Enhancing Customer Churn Prediction in Telecommunications: An Adaptive Ensemble Learning Approach," 2024. doi: 10.13140/RG.2.2.24573.78569.

[25] H. M. Jitendra Maan [1], "Customer Churn Prediction Model using Explainable Machine learning," *Int. J. Comput. Sci. Trends Technol.*, vol. 11, no. 1, pp. 33–38, 2023, doi: 10.48550/arXiv.2303.00960.

[26] S. Awasthi, "Customer Churn Prediction on E-Commerce Data using Stacking Classifier," Jul. 2022. doi: 10.36227/techrxiv.20291694.v1.

[27] A. M. Asfe, M. R. Rahman, and M. S. Hossain, "MNeuralTab: Integrating meta-modeling and neural networks for customer churn prediction in e-commerce," *Discov. Appl. Sci.*, vol. 7, no. 6, p. 569, May 2025, doi: 10.1007/s42452-025-07157-0.

[28] S. Kumar, S. Deep, and P. Kalra, "A Comprehensive Analysis of Machine Learning Techniques for Churn Prediction in E-Commerce: A Comparative Study," *Int. J. Comput. Trends Technol.*, vol. 72, no. 5, pp. 163–170, May 2024, doi: 10.14445/22312803/IJCTT-V72I5P119.