International Journal of Informatics and Information Systems
Vol. 9, No. 2, March 2026, pp. 357-370
ISSN 2579-7069

Predicting Customer Churn in E-Commerce Using Machine Learning: A Comprehensive Approach

Priya Sakthi Kumar[1,*], Vinoth Kumar P[2]

[1]AMET Deemed to be University, Kanathur, Chennai 603112, Tamil Nadu, India
[2]India Department of Electrical and Electronics Engineering, Sri Krishna College of Engineering and Technology, Tamil Nadu, India

(Received June 10, 2025; Revised October 15, 2025; Accepted January 30, 2026; Available online March 30, 2026)

Abstract

Customer churn prediction is a critical challenge in e-commerce, as retaining existing customers is often more cost-effective than acquiring new ones. This study evaluates the effectiveness of three machine learning algorithms Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) for predicting customer churn using a structured e-commerce dataset that integrates demographic, behavioral, engagement, and transactional features. A systematic preprocessing pipeline was implemented, including missing value imputation, categorical encoding, and feature scaling where appropriate. Model performance was assessed using accuracy, precision, recall, f1-score, and confusion Matrix analysis. The results reveal a clear performance gap between linear and ensemble-based approaches. LR achieved 77% accuracy but demonstrated limited recall for churned customers, indicating reduced sensitivity in identifying at-risk users. In contrast, RF and XGBoost both achieved 92% accuracy and substantially improved recall and f1-score for the churn class. Among the evaluated models, XGBoost showed the strongest overall performance, achieving the highest recall and lowest false negatives, making it particularly suitable for retention-focused applications. The findings confirm that ensemble and boosting-based models are more effective in capturing nonlinear interactions among multidimensional churn determinants in e-commerce environments. This study contributes empirical evidence supporting the adoption of advanced machine learning approaches for customer retention optimization and provides practical guidance for integrating churn prediction models into e-commerce platforms.

Keywords: Customer Churn Prediction, E-Commerce Analytics, Machine Learning, XGBoost, Customer Retention

1. Introduction

The e-commerce industry has witnessed remarkable growth in recent years, a trend significantly accelerated by the COVID-19 pandemic. As digital commerce became essential during the pandemic, e-commerce platforms quickly transitioned into critical infrastructure, supporting an increasing reliance on online shopping for the allocation of goods and services [1]. Prior to the pandemic, the e-commerce market was already projected to grow at an annual rate of approximately 9.8%, with substantial revenue increases expected through 2023 and beyond [1]. However, the pandemic acted as a catalyst for rapid consumer adoption across diverse markets, further embedding e-commerce as a fundamental part of the global economy [1]. Despite this growth, e-commerce businesses face significant challenges, particularly related to supply chain management, electronic fund transfers, and automated data collection systems. Additionally, regulatory frameworks have struggled to keep pace with the expansion of the sector, with governments implementing laws aimed at enhancing digital transparency, protecting consumer rights, and ensuring data regulation. These challenges highlight the dual nature of e-commerce growth: while presenting considerable commercial opportunities, the sector also confronts operational, regulatory, and sustainability challenges that require ongoing scholarly and policy attention [1].

One of the most pressing challenges for e-commerce businesses is customer churn the likelihood that a customer will discontinue their relationship with a platform. Churn poses a significant risk to long-term profitability, as customers at high risk of attrition are typically classified as low-loyalty customers, directly impacting revenue generation [2]. The

*Corresponding author: Priya Sakthi Kumar (priyasakthikumar@gmail.com)
DOI: https://doi.org/10.47738/ijiis.v9i2.300

financial implications of churn are profound, with studies showing that retaining existing customers is more cost-effective than acquiring new ones, underscoring the importance of customer loyalty in maintaining market share and reducing acquisition costs [3], [4].

Accurate churn prediction is a strategic imperative for e-commerce companies, as it enables proactive interventions, such as offering personalized discounts or improving customer service experiences, to at-risk customers before they disengage [2], [5]. Machine learning-based models for churn prediction have proven effective at identifying behavioral signals of customer disengagement, which are crucial for minimizing churn rates and improving long-term profitability [5]. Moreover, churn prediction is intrinsically linked to Customer Lifetime Value (CLV), as churn rates directly impact CLV calculations, which include factors like average order value and purchase frequency [6]. Therefore, minimizing churn enhances profitability by optimizing customer retention [6], [7], [8].

The complexity of churn is driven by various behavioral, demographic, and experiential factors that contribute to a customer's decision to disengage [6]. Unlike traditional retail, e-commerce platforms generate vast amounts of diverse data ranging from clickstream activity and transactional records to product interactions and service quality perceptions that must be systematically analyzed to identify churn-inducing signals [9]. This data complexity makes churn prediction a non-trivial task, requiring sophisticated machine learning techniques for accurate analysis [10], [11].

This research focuses on addressing the e-commerce industry's churn challenge by evaluating the effectiveness of machine learning models, specifically Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost), in predicting customer churn. Given the widespread use of these algorithms in classification tasks, this study aims to assess their performance and provide insights into the optimal model for churn prediction in e-commerce. The findings will contribute to the development of effective customer retention strategies, helping businesses minimize churn, enhance customer loyalty, and ultimately improve profitability.

2. Literature Review

2.1. Overview of Customer Churn in E-Commerce

Customer churn in e-commerce is shaped by a multidimensional interaction of demographic, behavioral, service-related, and engagement factors. Demographic characteristics such as age, gender, income level, and geographic location significantly influence purchasing patterns and churn propensity [6], [12]. When combined with transactional profiling, demographic segmentation helps identify churn-prone customer groups across retail and financial contexts [4]. Income level and regional differences further moderate engagement and purchasing behavior, particularly in cross-border e-commerce [13]. Integrated with behavioral data, these demographic variables enhance the predictive performance of machine learning-based churn models [6].

Behavioral determinants remain the most extensively studied predictors of churn. The traditional Recency Frequency Monetary (RFM) framework captures disengagement through inactivity, declining purchase frequency, and reduced spending [14]. However, recent research emphasizes that transactional signals alone are insufficient. Multi-behavior extensions incorporating clickstream activity, cart additions, login frequency, and session duration provide more granular early-warning indicators of disengagement [10], [15]. Social interaction behaviors, such as sharing and commenting, further enrich predictive capability, as declining engagement often precedes churn [9].

Service quality and broader engagement mechanisms also significantly influence churn outcomes. Logistics reliability, delivery performance, and website usability shape customer satisfaction and e-trust, which directly affect retention intentions [16]. Poor product information quality and negative service experiences increase platform switching behavior [17]. AI-driven personalization, marketing automation, and customer support systems further enhance engagement and retention by improving customer experience across the pre-purchase, purchase, and post-purchase stages [3]. Collectively, the literature underscores that churn in e-commerce is a complex phenomenon requiring integrated analytical approaches that combine demographic, behavioral, service quality, and engagement dimensions.

2.2. Churn Prediction Models in E-Commerce: Traditional Methods

Customer churn prediction in e-commerce is commonly formulated as a binary classification problem in which customers are labeled as churners or non-churners based on historical transactional and behavioral data. Traditional supervised learning algorithms, particularly LR and Decision Trees (DT), have long served as foundational approaches in churn prediction research due to their interpretability, computational efficiency, and suitability for structured datasets [18]. LR is valued for its probabilistic interpretation, enabling organizations to understand how specific features influence churn likelihood, and is often used as a baseline or meta-classifier in ensemble frameworks [19]. DTs provide intuitive rule-based structures that support managerial decision-making, although standalone trees are prone to overfitting and limited generalization [19].

To overcome these limitations, ensemble methods such as RF and gradient boosting models have gained prominence. RF, which aggregates multiple decision trees using bagging, is widely reported as a robust baseline capable of handling nonlinear feature interactions and heterogeneous data [18]. More recently, XGBoost has emerged as a leading gradient boosting algorithm, frequently achieving superior accuracy, recall, and AUC in large-scale or imbalanced churn datasets [7], [8], [9]. By sequentially correcting prior errors, XGBoost effectively captures complex behavioral patterns and enhances predictive stability.

Despite the strong performance of boosting approaches, the literature consistently emphasizes that no single model universally dominates across all contexts. Model effectiveness remains dependent on dataset characteristics, feature richness, class imbalance severity, and evaluation metrics [19]. Consequently, hybrid and stacking strategies combining LR, RF, and XGBoost are increasingly recommended to leverage complementary strengths, balancing interpretability with predictive power [7]. Overall, while traditional models remain essential as interpretable baselines, ensemble and boosting methods represent the prevailing direction for achieving state-of-the-art churn prediction performance in e-commerce.

2.3. Previous Studies in Churn Prediction

The literature on Customer Churn Prediction (CCP) across e-commerce, telecommunications, and financial services demonstrates a sustained reliance on machine learning techniques to identify at-risk customers and enhance retention strategies. Early research framed churn prediction as a binary classification task, primarily employing traditional algorithms such as Logistic Regression LR and DT [18], [19]. Over time, the field has evolved toward ensemble and boosting approaches, accompanied by greater emphasis on feature engineering, class imbalance handling, and rigorous evaluation frameworks [20]. This evolution reflects a transition from simple baseline models to more integrated and high-performance churn analytics systems.

Across domains, LR remains a widely used interpretable baseline, particularly in structured or smaller datasets where explainability is essential. DTs provide intuitive rule-based classification but are often incorporated into ensemble structures due to overfitting risks [19]. RF, as a bagging-based ensemble of decision trees, has demonstrated strong generalization performance and is frequently reported as a robust baseline [21]. More recently, gradient boosting methods especially XGBoost have emerged as leading performers, achieving superior accuracy and AUC in complex or large-scale churn datasets [9]. Hybrid and stacking approaches combining LR, RF, and XGBoost are increasingly adopted to leverage complementary strengths [7].

Comparative studies consistently report that ensemble methods, particularly XGBoost, outperform standalone traditional classifiers in terms of accuracy, precision, recall, F1-score, and AUC [19], [22]. However, performance remains dataset-dependent; in moderately sized or less complex datasets, LR or RF may remain competitive. Preprocessing strategies such as SMOTE and hybrid resampling significantly influence minority-class detection and overall robustness [23]. Moreover, feature engineering plays a central role, with extended RFM models, clickstream signals, and engagement metrics enhancing predictive power [15]. Despite methodological advances, the literature emphasizes that no single model universally dominates; effective churn prediction requires dataset-aware model selection, balanced evaluation metrics, and integrated feature design tailored to domain-specific customer behavior patterns.

2.4. Gaps in the Literature on Churn Prediction in E-Commerce

Despite the substantial body of research on CCP, two key gaps remain prominent in the e-commerce domain. First, relatively few studies conduct rigorously controlled comparisons of multiple machine learning algorithms within a unified experimental framework. Although LR, RF, and XGBoost are widely recognized as effective models across sectors, systematic benchmarking under identical preprocessing conditions, sampling strategies, and evaluation metrics within a single e-commerce dataset is still limited [18], [19]. Many studies emphasize a single high-performing model often XGBoost without comprehensive side-by-side comparisons against traditional baselines, making it difficult to draw generalizable conclusions about relative performance. Differences in imbalance handling (e.g., SMOTE, ENN), feature engineering, and evaluation criteria further complicate cross-study comparability [23].

A second major gap concerns the limited systematic investigation of feature contributions. While prior studies incorporate RFM variables, multi-behavior signals (clicks, cart additions, favorites), and session-level metrics, few employ structured ablation experiments to quantify the incremental predictive value of each feature family within a consistent modeling pipeline [14], [15]. As a result, practitioners lack clear guidance on which behavioral or engagement signals should be prioritized, particularly under resource constraints. Moreover, interactions between feature types and model choice such as whether session-based metrics disproportionately enhance boosting models relative to linear classifiers remain underexplored [20].

These limitations are further compounded by contextual heterogeneity across e-commerce settings. In cross-border or omni-channel environments, the importance of engagement, operational quality, and session-level behavior may vary significantly [13], [15]. Without controlled experimental designs and standardized benchmarking protocols, it is difficult to determine whether performance differences arise from algorithmic superiority or dataset-specific characteristics [19]. Addressing these gaps requires a structured research agenda emphasizing rigorous algorithm comparison and systematic feature ablation within a shared e-commerce dataset. Such an approach employing consistent data partitions, standardized imbalance handling, and comprehensive evaluation metrics (accuracy, precision, recall, F1, AUC) would provide clearer evidence for model selection and feature prioritization, thereby strengthening the methodological foundation of e-commerce churn analytics.

3. Methodology

3.1. Dataset Collection

This study utilizes a structured e-commerce customer dataset designed for churn prediction analysis. Each record represents an individual customer and includes a binary target variable (Churned) indicating whether the customer has discontinued engagement with the platform. The dataset integrates demographic, behavioral, engagement, and financial attributes to support supervised machine learning modeling.

To provide a clearer overview of the dataset structure, the feature categories and their corresponding attributes are summarized in Table 1.

Table 1. Feature Categories and Attributes Used for Churn Prediction

Feature Category	Attributes	Description
Demographic Features	Age, Gender, Country, City	Customer socio-demographic characteristics used for segmentation and profiling
Tenure & Usage Metrics	Membership_Years, Login_Frequency, Session_Duration_Avg, Pages_Per_Session	Indicators of customer tenure and platform interaction intensity
Behavioral Features	Cart_Abandonment_Rate, Wishlist_Items, Product_Reviews_Written, Social_Media_Engagement_Score, Mobile_App_Usage	Behavioral signals reflecting purchase intent, engagement level, and digital interaction patterns
Communication & Service Interaction	Email_Open_Rate, Customer_Service_Calls	Measures of responsiveness to marketing communication and post-purchase service interactions
Financial & Transactional Features	Payment_Method_Diversity, Lifetime_Value, Credit_Balance	Indicators of customer economic value and purchasing capacity
Temporal Attribute	Signup_Quarter	Customer acquisition timing information
Target Variable	Churned	Binary label (1 = churned, 0 = retained)

As shown in Table 1, the dataset is multidimensional and captures diverse aspects of customer behavior and value. The integration of demographic, behavioral, financial, and engagement-related attributes enable comprehensive churn

 ISSN 2579-7069

modeling and facilitates comparative evaluation of machine learning algorithms. Moreover, the structured categorization supports subsequent feature importance analysis to determine which feature families contribute most significantly to churn prediction in the e-commerce context.

3.2. Data Preprocessing

Prior to model development, a systematic preprocessing procedure was conducted to ensure data quality, consistency, and suitability for machine learning algorithms. Missing values were handled using imputation techniques to preserve dataset integrity and avoid unnecessary data loss. For numerical variables such as session duration, cart abandonment rate, and lifetime value, missing values were imputed using the median, as it is robust to outliers and maintains the underlying distribution. For categorical variables such as gender, country, city, and signup quarter, missing entries were replaced using the mode (most frequent category), ensuring logical consistency without introducing artificial bias.

Because machine learning algorithms require numerical input, categorical variables were transformed into numeric representations using Label Encoding. Attributes including Gender, Country, City, and Signup_Quarter were converted into integer-encoded values. LabelEncoder was selected due to its computational simplicity and compatibility with tree-based algorithms such as Random Forest and XGBoost, which are not sensitive to ordinal encoding effects. For the Logistic Regression model, the encoded categorical variables were treated as nominal predictors within the modeling framework.

Feature scaling was applied selectively depending on the algorithmic requirements. Numerical features were standardized using a normalization technique (e.g., StandardScaler) for Logistic Regression to ensure comparable feature magnitudes and to improve convergence stability. In contrast, scaling was not strictly required for tree-based models such as RF and XGBoost, as these algorithms rely on split-based decision rules rather than distance-based calculations. These preprocessing steps ensured that the dataset was clean, consistently formatted, and suitable for fair comparative evaluation across the selected machine learning models.

3.3. Model Selection

Three machine learning algorithms were selected to evaluate churn prediction performance, representing linear, bagging-based ensemble, and boosting-based ensemble paradigms. The selection of LR, RF, and XGBoost is grounded in prior literature that consistently positions these models as foundational, robust, and state-of-the-art approaches in customer churn prediction [19].

LR was employed as a baseline linear classification model. As a generalized linear model, LR estimates the probability of churn using the logistic (sigmoid) function applied to a linear combination of input features:

$$P(y = 1 \mid X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)}} \quad (1)$$

where $\beta_0$is the intercept, $\beta_i$are model coefficients, and $x_i$are feature values. LR is consistently described in the literature as a simple, interpretable baseline classifier valued for transparency and computational efficiency [24]. It is frequently used as a reference model against which more complex algorithms are compared, and often serves as a meta-learner in stacking ensembles combining RF and XGBoost [7]. This role is particularly important in contexts where interpretability is prioritized, such as managerial decision-making or regulatory environments. However, LR may underperform in highly nonlinear or severely imbalanced datasets, where ensemble or boosting methods typically yield superior predictive performance [7], [24]. Despite this limitation, LR remains a strong, interpretable benchmark and a viable component in hybrid configurations.

RF was selected as a robust ensemble learning method based on bootstrap aggregating (bagging). It constructs multiple decision trees $T_1, T_2, \ldots, T_K$ and aggregates their predictions through majority voting:

$$\hat{y} = \text{mode}\{T_1(X), T_2(X), \ldots, T_K(X)\} \quad (2)$$

ISSN 2579-7069

Each tree is typically built using impurity measures such as Gini Index:

$$Gini = 1 - \sum_{i=1}^{C} p_i^2 \tag{3}$$

where p_i represents the proportion of class $i$in a node. RF is widely cited as a robust baseline for churn prediction due to its strong generalization ability and resistance to overfitting through bagging [19], [24]. Empirical studies frequently report high accuracy and F1 scores for RF across churn contexts, and it is often integrated into hybrid ensembles with LR and XGBoost to leverage complementary strengths [24]. Nevertheless, RF performance depends on hyperparameter tuning (e.g., number of trees, maximum depth) and data characteristics. In moderately nonlinear datasets, RF may approach XGBoost performance, especially when dataset size is not extremely large [19].

XGBoost was chosen as a leading gradient-boosting framework designed for high predictive accuracy. Unlike RF, which builds trees independently, XGBoost constructs trees sequentially, where each new tree corrects errors from previous iterations:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \tag{4}$$

The objective function minimized at each iteration is:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t)}) + \sum_{t=1}^{T} \Omega(f_t) \tag{5}$$

with regularization term:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{6}$$

where T is the number of leaves, w_j are leaf weights, and γ, λ are regularization parameters. XGBoost is repeatedly highlighted as a state-of-the-art model in churn prediction literature, often outperforming LR and RF in large, complex, or imbalanced datasets [19]. It is particularly effective when feature engineering (e.g., RFM extensions, temporal signals) and class imbalance handling are applied [24]. However, in smaller or linearly separable datasets, LR or RF may achieve comparable performance without the added computational complexity [7], [24].

Based on the synthesis of prior studies, a structured model selection strategy is recommended. First, LR should be implemented to establish a transparent and interpretable baseline. Second, RF should be included as a robust non-linear ensemble baseline capable of handling mixed feature types with minimal preprocessing. Third, XGBoost should be evaluated as the primary high-performance comparator, particularly for larger or more complex datasets. Finally, hybrid stacking approaches combining LR, RF, and XGBoost with LR as a meta-learner may be considered to exploit complementary strengths and enhance overall predictive performance [7], [24]. Importantly, data conditioning practices such as class imbalance handling (e.g., SMOTE) and feature engineering often influence model performance as strongly as algorithm choice itself [24], underscoring the need for careful experimental design in churn prediction studies.

3.4. Model Evaluation Metrics

Model performance was evaluated using multiple classification metrics to ensure a comprehensive assessment, particularly given the potential class imbalance in churn datasets. Relying solely on accuracy can be misleading when churned customers represent a minority class; therefore, additional metrics such as Precision, Recall, and F1-Score were included to provide a balanced evaluation.

Accuracy measures the proportion of correctly classified instances among all predictions and is defined as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{7}$$

where TP (True Positive) represents correctly predicted churned customers, TN (True Negative) represents correctly predicted retained customers, FP (False Positive) represents customers incorrectly predicted as churned, and FN (False Negative) represents churned customers incorrectly predicted as retained.

Precision evaluates the correctness of positive predictions and is particularly important when false alarms (incorrectly labeling customers as churned) carry business implications:

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

A high precision value indicates that when the model predicts churn, it is likely to be correct.

Recall (also called Sensitivity or True Positive Rate) measures the model's ability to correctly identify actual churned customers:

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

F1-Score provides a harmonic mean of Precision and Recall, balancing both metrics:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (10)$$

The f1-score is particularly useful when class imbalance exists and when both false positives and false negatives are important.

In addition to numerical metrics, the confusion matrix was used to provide a detailed breakdown of classification outcomes. The confusion matrix is structured as:

$$\begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix} \quad (11)$$

This matrix enables visual inspection of how predictions are distributed across actual and predicted classes. It supports deeper interpretation of model behavior, including whether the model is biased toward predicting the majority class or effectively capturing churned customers.

By combining accuracy, precision, recall, f1-score, and confusion matrix analysis, the evaluation framework ensures a robust and balanced comparison across Logistic Regression, Random Forest, and XGBoost models in the e-commerce churn prediction task.

4. Results

4.1. Model Performance

To provide a structured evaluation of the proposed models, the performance of LR, RF, and XGBoost is presented individually. Each model is analyzed using Accuracy, Precision, Recall, and F1-Score to ensure a balanced assessment, particularly for the churned customer class (Class 1), which is critical for business decision-making. The detailed evaluation results for each model are shown in the following tables. The performance metrics for the LR model are summarized in Table 2.

Table 2. LR Evaluation Metrics

Metric	Class 0 (Retained)	Class 1 (Churned)	Overall
Precision	0.80	0.67	–
Recall	0.92	0.41	–
F1-Score	0.85	0.51	–
Accuracy	–	–	0.77
Macro Avg F1	–	–	0.68
Weighted Avg F1	–	–	0.75

As shown in Table 2, LR achieved an overall accuracy of 77%. The model demonstrates strong performance in identifying retained customers (recall = 0.92), but its ability to detect churned customers is considerably weaker (recall

= 0.41). This indicates a relatively high number of false negatives, meaning many customers who actually churned were not correctly identified. Although LR offers interpretability and computational efficiency, its linear structure limits its ability to model complex nonlinear interactions within customer behavioral data. The evaluation results for the RF model are presented in Table 3.

Table 3. RF Evaluation Metrics

Metric	Class 0 (Retained)	Class 1 (Churned)	Overall
Precision	0.91	0.92	–
Recall	0.97	0.77	–
F1-Score	0.94	0.84	–
Accuracy	–	–	0.92
Macro Avg F1	–	–	0.89
Weighted Avg F1	–	–	0.91

As illustrated in Table 3, RF achieved a substantial improvement over Logistic Regression, reaching 92% overall accuracy. The recall for churned customers increased significantly to 0.77, indicating a strong reduction in false negatives. The balanced macro-average F1-score (0.89) confirms that the model performs consistently across both retained and churned classes. The ensemble nature of RF enables it to capture nonlinear relationships and feature interactions more effectively than linear models. The performance metrics for the XGBoost model are displayed in Table 4.

Table 4. XGBoost Evaluation Metrics

Metric	Class 0 (Retained)	Class 1 (Churned)	Overall
Precision	0.92	0.91	–
Recall	0.97	0.78	–
F1-Score	0.94	0.84	–
Accuracy	–	–	0.92
Macro Avg F1	–	–	0.89
Weighted Avg F1	–	–	0.91

As shown in Table 4, XGBoost achieved performance comparable to Random Forest, with 92% overall accuracy. Notably, it achieved the highest recall for churned customers (0.78), slightly outperforming Random Forest in identifying at-risk customers. The model maintains high precision (0.91) and a strong F1-score (0.84) for the churn class, indicating an effective balance between detecting true churn cases and minimizing false alarms. The gradient boosting mechanism enables XGBoost to sequentially optimize prediction errors and handle complex data patterns efficiently.

Overall, the results demonstrate that ensemble-based approaches (RF and XGBoost) significantly outperform the linear baseline (LR), particularly in detecting churned customers. While LR remains valuable for interpretability and baseline benchmarking, XGBoost and RF provide superior predictive performance and are better suited for practical churn management strategies in e-commerce environments.

4.2. Confusion Matrix Analysis

To further understand the classification behavior of each model, confusion matrices were analyzed to examine the distribution of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). This analysis provides deeper insight into how well each model distinguishes between retained (Active) and churned customers. The confusion matrix for LR is illustrated in Figure 1.

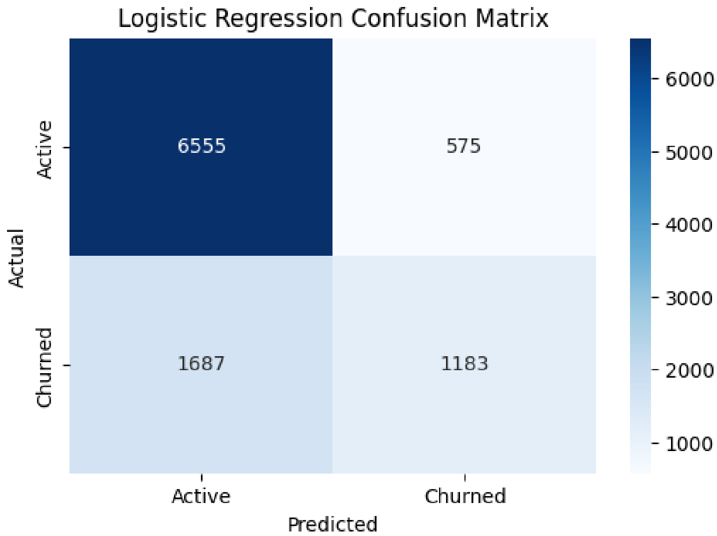

Figure 1. Confusion Matrix of Logistic Regression Model for Customer Churn Prediction

The model correctly identified a large number of retained customers (TN), but it misclassified a substantial number of churned customers as retained (high FN = 1,687). This explains the relatively low recall (0.41) for churned customers observed earlier. In practical terms, many at-risk customers would not be flagged for retention intervention, limiting the model's business effectiveness. The confusion matrix for RF is presented in Figure 2.

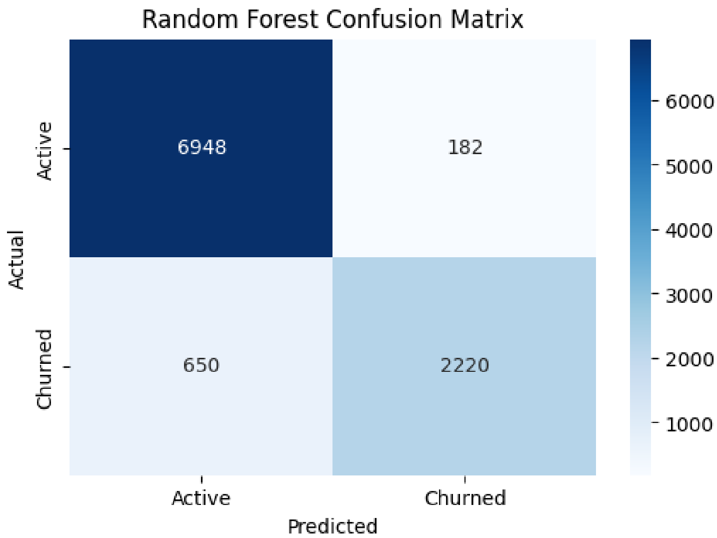

Figure 2. Confusion Matrix of Random Forest Model for Customer Churn Prediction

RF significantly reduced both false positives and false negatives compared to LR. The sharp decline in FN (from 1,687 to 650) indicates much better detection of churned customers. At the same time, the relatively low FP count (182) shows that the model does not excessively misclassify retained customers as churned. This balance explains the strong F1-score and overall accuracy of 92%. The confusion matrix for XGBoost is shown in Figure 3.

ISSN 2579-7069

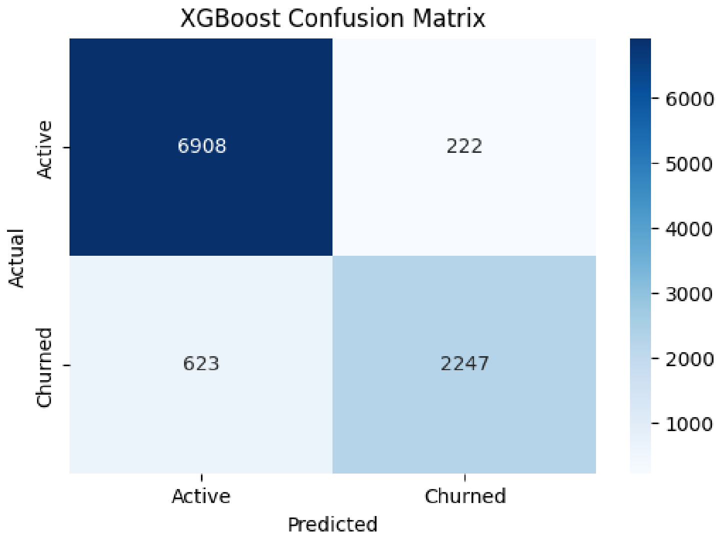

Figure 3. Confusion Matrix of XGBoost Model for Customer Churn Prediction

XGBoost achieved the lowest number of FN (623), slightly outperforming RF in identifying churned customers. Although its FP (222) are marginally higher than RF, the model maintains a strong balance between sensitivity and precision. The improved TP value (2,247) reflects its superior recall (0.78) for churn detection.

Comparatively, LR demonstrates strong performance in predicting retained customers but struggles in identifying churn cases due to high false negatives. RF and XGBoost substantially improve churn detection performance. Among the three models, XGBoost achieves the highest number of correctly identified churned customers (TP), while RF yields the lowest false positive rate.

From a business perspective, reducing false negatives (missed churners) is often more critical than minimizing false positives. Therefore, XGBoost offers the most advantageous trade-off for customer retention strategies, followed closely by RF.

4.3. Comparative Analysis

To provide a consolidated comparison across the three models, their key evaluation metrics are summarized in Table 5. The comparison focuses particularly on the churned class (Class 1), as correctly identifying at-risk customers is the primary objective of churn prediction.

Table 5. Comparative Performance of LR, RF, and XGBoost

Model	Accuracy	Precision (Churn)	Recall (Churn)	F1-Score (Churn)	False Negatives	False Positives
LR	0.77	0.67	0.41	0.51	1,687	575
RF	0.92	0.92	0.77	0.84	650	182
XGBoost	0.92	0.91	0.78	0.84	623	222

As shown in Table 5, ensemble-based models (RF and XGBoost) significantly outperform LR across all key metrics. Logistic Regression achieved only 77% accuracy and exhibited low recall (0.41) for churned customers, indicating that a substantial number of at-risk customers were not detected (high FN = 1,687). While LR provides interpretability advantages, its predictive power is limited in this dataset.

RF achieved strong overall performance with 92% accuracy and high precision (0.92) and recall (0.77) for churned customers. It recorded the lowest number of FP (182), suggesting that it is slightly more conservative in labeling customers as churned.

XGBoost achieved the highest recall for churned customers (0.78) and the lowest number of FN (623), meaning it correctly identified the greatest number of churned customers. Although its FP (222) are slightly higher than RF, the difference is relatively small. Given that reducing FN is typically more critical in churn management since undetected churners represent lost revenue opportunities XGBoost provides the most favorable trade-off between sensitivity and overall predictive performance.

Overall, while LR serves as a useful baseline, XGBoost emerges as the best-performing model for churn prediction in this study, closely followed by RF. The results confirm that boosting and ensemble approaches are more effective than linear models in capturing complex behavioral patterns in e-commerce churn datasets.

5. Discussion

The findings of this study carry important practical implications for e-commerce businesses seeking to enhance customer retention strategies. The superior performance of ensemble models particularly XGBoost and RF confirms that nonlinear, interaction-aware algorithms are better suited for modeling the multidimensional determinants of churn behavior, including demographic, behavioral, service-quality, and engagement-related factors [18], [19]. Models with higher recall for churned customers, such as XGBoost, are especially valuable in practice because they reduce false negatives and enable earlier identification of at-risk customers, thereby supporting timely retention interventions. This aligns with prior research emphasizing the strategic importance of predictive analytics in Customer Relationship Management (CRM) systems for proactive churn mitigation [3].

The results also reinforce the critical role of multidimensional behavioral and engagement signals in improving predictive performance. Consistent with the extended RFM framework and multi-behavior modeling approaches, variables such as login frequency, session duration, cart abandonment rate, and transactional value contribute significantly to churn detection [14], [15]. E-commerce platforms should therefore prioritize comprehensive data collection pipelines that integrate transactional history, session-level activity, and service interaction metrics. Embedding churn prediction outputs into recommendation systems and AI-driven marketing automation platforms can enhance personalization, reduce disengagement, and improve long-term customer loyalty [25]. In this way, predictive modeling becomes an operational component of digital customer experience optimization rather than a purely analytical exercise.

Despite these contributions, several limitations must be acknowledged. First, the dataset does not incorporate advanced temporal sequences or sentiment-based features derived from customer reviews, which have been shown to provide early indicators of dissatisfaction and churn [4]. Second, although LR, RF, and XGBoost represent foundational and state-of-the-art approaches in churn prediction [9], [18], more advanced deep learning architectures or hybrid stacking frameworks may further enhance predictive robustness [7]. Additionally, the literature emphasizes that model performance is highly dependent on dataset characteristics, feature richness, and imbalance handling strategies [19], suggesting that findings may vary under different data conditions.

Future research should therefore conduct structured feature ablation analyses to quantify the incremental contribution of demographic, behavioral, service-quality, and engagement-related feature families, addressing the feature-contribution gap highlighted in prior studies [20]. Incorporating additional data modalities such as review sentiment, real-time clickstream data, and cross-channel engagement metrics could strengthen predictive accuracy and generalizability across different e-commerce contexts [13], [15]. Furthermore, exploring advanced architectures, including sequential deep learning models, may better capture dynamic customer behavior patterns over time. Collectively, these directions would advance the development of more integrated, context-aware, and practically deployable churn prediction frameworks in digital commerce ecosystems.

6. Conclusion

This study evaluated the performance of LR, RF, and XGBoost for predicting customer churn in an e-commerce dataset incorporating demographic, behavioral, engagement, and transactional features. The results show that ensemble-based models significantly outperform the linear baseline. While Logistic Regression achieved moderate accuracy (77%) and provided interpretability advantages, RF and XGBoost both reached 92% accuracy. Among them, XGBoost demonstrated the strongest ability to identify churned customers, achieving the highest recall and lowest false negatives. These findings confirm that nonlinear ensemble methods are better suited to capturing the complex patterns underlying e-commerce churn behavior.

This research contributes to the churn prediction literature by providing a controlled comparative evaluation of linear and ensemble models within a unified experimental framework. The study reinforces prior findings that boosting-based approaches often outperform traditional classifiers in structured churn datasets, while also highlighting the continued relevance of LR as an interpretable benchmark. Moreover, the integration of multidimensional features supports the view that churn is influenced by combined demographic, behavioral, and engagement factors.

Overall, the study emphasizes the importance of machine learning-driven churn analytics for improving customer retention and strategic decision-making in e-commerce. Implementing high-performing predictive models such as XGBoost can enable proactive retention initiatives, enhance personalization, and support sustainable business growth in increasingly competitive digital markets.

7. Declarations

6.1. Author Contributions

Author Contributions: Conceptualization P.S.K. and V.K.P.; Methodology, P.S.K. and V.K.P.; Software, P.S.K.; Validation, P.S.K.; Formal Analysis, P.S.K.; Investigation, V.K.P.; Resources, P.S.K.; Data Curation, V.K.P.; Writing Original Draft Preparation, P.S.K.; Writing Review and Editing, P.S.K. and V.K.P.; Visualization, V.K.P. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. M. Branch, “Union Exemption: Nonprofit Work and the Boundaries of the Commercial Economy, 1951–1976,” *Mod. Am. Hist*., vol. 8, no. 1, pp. 73–93, 2025, doi: 10.1017/mah.2024.42.

[2] S. Roosta, S. J. Sadjadi, and A. Makui, “Predicting Customer Loyalty in Omnichannel Retailing Using Purchase Behavior, Socio-Cultural Factors, And Learning Techniques,” PL*oS One*, vol. 20, no. 8, pp. e0330338, 2025, doi: 10.1371/journal.pone.0330338.

[3] P. Chugh and V. Jain, “Artificial Intelligence (AI) Empowerment in E-Commerce: A Bibliometric Voyage,” *NMIMS Manag. Rev*., vol. 32, no. 3, pp. 159–173, 2024, doi: 10.1177/09711023241303621.

[4] C. Singh, M. K. Dash, R. Sahu, and A. Kumar, “Artificial Intelligence in Customer Retention: A Bibliometric Analysis and Future Research Framework,” *Kybernetes*, vol. 53, no. 11, pp. 4863–4888, 2024, doi: 10.1108/K-02-2023-0245.

[5] N.-T. Lee, H.-C. Lee, J. Hsin, and S. Fang, “Prediction of Customer Behavior Changing via a Hybrid Approach,” *IEEE Open J. Comput. Soc*., vol. 5, no. 1, pp. 27–38, 2024, doi: 10.1109/OJCS.2023.3336904.

ISSN 2579-7069

[6] K. V Ramana, M. Nalini, D. C. Devi, B. Maheswari, and S. Subramanian, "Customer churn prediction model in enterprises using machine learning," *Int. J. Adv. Technol. Eng. Explor.*, vol. 11, no. 110, pp. 94–107, 2024, doi: 10.19101/IJATEE.2023.10101581.

[7] N. Ahmad, M. J. Awan, H. Nobanee, A. M. Zain, A. Naseem, and A. Mahmoud, "Customer Personality Analysis for Churn Prediction Using Hybrid Ensemble Models and Class Balancing Techniques," *IEEE Access*, vol. 12, no. 1, pp. 1865–1879, 2024, doi: 10.1109/ACCESS.2023.3334641.

[8] I. AlShourbaji, N. Helian, Y. Sun, A. G. Hussien, L. Abualigah, and B. Elnaim, "An efficient Churn Prediction Model Using Gradient Boosting Machine and Metaheuristic Optimization," *Sci. Rep.*, vol. 13, no. 1, pp. 14441, 2023, doi: 10.1038/s41598-023-41093-6.

[9] 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.

[10] M. A. Al Rahib, N. Saha, R. Mia, and A. Sattar, "Customer Data Prediction and Analysis in E-Commerce Using Machine Learning," *Bull. Electr. Eng. Informatics*, vol. 13, no. 4, pp. 2624–2633, 2024, doi: 10.11591/eei.v13i4.6420.

[11] E. M. Elgohary, M. Galal, A. Mosa, and G. A. Elshabrawy, "Smart Evaluation for Deep Learning Model: Churn Prediction as a Product Case Study," *Bull. Electr. Eng. Informatics*, vol. 12, no. 2, pp. 1219–1225, 2023, doi: 10.11591/eei.v12i2.4180.

[12] K. Dhanushkodi, A. Bala, N. Kodipyaka, and V. Shreyas, "Customer Behavior Analysis and Predictive Modeling in Supermarket Retail: A Comprehensive Data Mining Approach," *IEEE Access*, vol. 13, no. 1, pp. 2945–2957, 2025, doi: 10.1109/ACCESS.2024.3407151.

[13] F. Le and J. Zhai, "Research on Cross-Border E-Commerce Customer Churn Prediction Based on Enhanced Xgboost Algorithm with Temporal-Spatial Features," *J. Comput. Methods Sci. Eng.*, vol. 25, no. 5, pp. 4407–4418, 2025, doi: 10.1177/14727978251337888.

[14] Y. Sun, H. Liu, and Y. Gao, "Research on Customer Lifetime Value Based On Machine Learning Algorithms And Customer Relationship Management Analysis Model," Heliyon, vol. 9, no. 2, pp. e13384, 2023, doi: 10.1016/j.heliyon.2023.e13384.

[15] Q. Zhao, Z. Zhao, L. Yang, L. Hong, and W. Han, "Research on Customer Retention Prediction Model of VOD Platform Based on Machine Learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 4, pp. 236–244, 2023, doi: 10.14569/IJACSA.2023.0140427.

[16] W. Rachbini, "The Impact of Price Transparency, Safe Transactions, and Delivery Performance on E-commerce Performance in Indonesian Online Supermarkets," *Eur. J. Manag. Issues*, vol. 31, no. 3, pp. 132–141, 2023, doi: 10.15421/192311.

[17] S. Amsl, I. Watson, C. Teller, and S. Wood, "Presenting products on websites – the importance of information quality criteria for online shoppers," *Int. J. Retail Distrib. Manag.*, vol. 51, no. 9/10, pp. 1213–1238, 2023, doi: 10.1108/IJRDM-04-2023-0266.

[18] P. S. P. Kumar and C. K. Ramu, "Two-Dimensional Customer Segmentation for Digital Custom Coupon Issuance Approach," *Int. J. Multidiscip. Res. Sci. Eng. Technol.*, vol. 7, no. 05, pp. 10042–10052, 2024, doi: 10.15680/IJMRSET.2024.0705076.

[19] L. Geiler, S. Affeldt, and M. Nadif, "A Survey on Machine Learning Methods for Churn Prediction," *Int. J. Data Sci. Anal.,* vol. 14, no. 3, pp. 217–242, 2022, doi: 10.1007/s41060-022-00312-5.

[20] U. Gani Joy, K. E. Hoque, M. Nazim Uddin, L. Chowdhury, and S.-B. Park, "A Big Data-Driven Hybrid Model for Enhancing Streaming Service Customer Retention Through Churn Prediction Integrated with Explainable AI," *IEEE Access*, vol. 12, no. 1, pp. 69130–69150, 2024, doi: 10.1109/ACCESS.2024.3401247.

[21] K. Şahinbaş, "Predicting Order Cancellations for E-Commerce Domain: A Proposed Model Based on Retailing Experience," *İnsan Ve Toplum Bilim. Araştırmaları Derg*., vol. 11, no. 3, pp. 1493–1514, 2022, doi: 10.15869/itobiad.1127578.

[22] B. Durkaya Kurtcan and T. Ozcan, "Predicting Customer Churn Using Grey Wolf Optimization-Based Support Vector Machine with Principal Component Analysis," *J. Forecast*., vol. 42, no. 6, pp. 1329–1340, 2023, doi: 10.1002/for.2960.

[23] J.-X. Ong, G.-K. Tong, K. Khor, and S.-C. Haw, "Enhancing Customer Churn Prediction with Resampling: A Comparative Study," *TEM J*., vol. 13, no. 3, pp. 1927–1936, 2024, doi: 10.18421/TEM133-20.

[24] A. Dhini and M. Fauzan, "Predicting Customer Churn using Ensemble Learning: Case Study of a Fixed Broadband Company," *Int. J. Technol*., vol. 12, no. 5, 1030, 2021, doi: 10.14716/ijtech.v12i5.5223.

[25] R. E. Bawack, S. F. Wamba, K. D. A. Carillo, and S. Akter, "Artificial intelligence in E-Commerce: A Bibliometric Study and Literature Review," *Electron. Mark*., vol. 32, no. 1, pp. 297–338, 2022, doi: 10.1007/s12525-022-00537-z.