An Intelligent CRM System for Early Identification of Customer Attrition in E-Commerce

Dr Chintal Kumar Patel
Associate Professor, CSE
Geetanjali Institute of Technical Studies, Dabok, Rajasthan
chintal.patel@gits.ac.in

Abstract: *The prediction of customer attrition in the banking industry has gained significant attention due to the proliferation of services and the need to retain customers in today's competitive market, particularly when customers abandon banking products and services. Customer churn, defined as the termination of a customer's relationship with a bank, has become a critical concern. This study proposes an intelligent machine learning framework for customer attrition prediction using the E-Commerce Customer Churn Dataset. Two ensemble learning algorithms, XGBoost and Random Forest, are implemented and evaluated using accuracy, precision, recall, and F1-score. Experimental results demonstrate that the proposed XGBoost model achieves the highest accuracy of 98.9%, with 96.8% precision, 94.7% recall, and 95.6% F1-score, while the proposed Random Forest model achieves 98.4% accuracy, 97.1% precision, 90.9% recall, and 94.5% F1-score. Furthermore, the proposed models outperform existing machine learning approaches, including Logistic Regression, Stochastic Gradient Boosting, Naïve Bayes, Support Vector Machine, and Artificial Neural Network. The proposed intelligent CRM framework provides accurate customer attrition prediction, enabling e-commerce organizations to identify at-risk customers early, improve retention strategies, enhance customer loyalty, and support data-driven business decision-making.*

Keywords: Customer Churn Prediction, E-Commerce Analytics, Machine Learning, Customer Relationship Management, Early Detection.

I. INTRODUCTION

The rapid growth of e-commerce has transformed the global retail landscape by enabling businesses to deliver products and services through digital platforms[1]. The widespread adoption of smartphones, high-speed internet, and secure online payment systems has significantly changed consumer purchasing behavior, making online shopping more convenient and accessible[2][3][4]. E-commerce platforms generate enormous volumes of customer data through transactions[5][6], browsing activities, reviews, and digital interactions, providing businesses with valuable opportunities to understand customer preferences and improve decision-making[7][8]. However, increasing competition in the digital marketplace has made customer retention a critical challenge, as acquiring new customers is considerably more expensive than retaining existing ones. Therefore, organizations are increasingly focusing on long-term relationship management to sustain profitability and competitive advantage[9][10].

Customer Relationship Management (CRM) has emerged as a strategic approach for managing customer interactions, improving customer satisfaction, and strengthening long-term relationships[11][12][13]. Modern CRM systems integrate customer information from multiple channels into a centralized platform, enabling organizations to monitor customer behavior, personalize services, and support data-driven business decisions. In the e-commerce environment, effective CRM practices facilitate targeted marketing[14], efficient customer service, and improved customer engagement[15][16][17]. However, traditional CRM systems that primarily store and manage customer information are no longer sufficient to address the growing complexity of customer behavior across multiple digital touchpoints[18][19][20].

One of the major challenges faced by e-commerce businesses is customer attrition (customer churn), where customers discontinue their purchases or abandon a platform over time[21]. Customer churn directly impacts organizational revenue, customer lifetime value, and brand loyalty, making its early identification essential for sustainable business growth[22][23]. Conventional churn prediction methods mainly rely on historical reports, predefined rules, and manual analysis, providing only retrospective insights and failing to identify potential churn proactively[24]. Moreover, these approaches are unable to efficiently process the large-scale, high-dimensional, and continuously evolving customer data generated from digital platforms[25][26]. Consequently, businesses require intelligent predictive systems capable of identifying at-risk customers before they disengage, enabling timely retention strategies and personalized interventions[27][28][29].

Recent advances in Machine Learning (ML) and Deep Learning (DL) have significantly enhanced intelligent CRM systems by enabling automated analysis of complex customer data and accurate prediction of customer behavior. ML and DL algorithms learn hidden patterns from demographic information, transaction history, browsing behavior, purchase frequency, and customer engagement metrics to predict churn with high accuracy. Compared with conventional statistical approaches, these intelligent models provide superior predictive performance, adaptability, and scalability for large e-commerce datasets[30][31].

A. Motivation and Contribution

The drive behind this research is the growing problem of customer churn in the burgeoning e-commerce industry, where maintaining customer loyalty is critical to long-term business success. Traditional CRM systems and conventional prediction methods are often unable to effectively analyze large-scale and complex customer behavior data. Therefore, there is a strong need for an intelligent machine learning-based framework that can accurately identify potential churn customers at an early stage. Such a system enables proactive retention strategies, enhances customer satisfaction, and improves overall business performance. This research offers several key contributions as listed below:

Developed an intelligent Customer Relationship Management (CRM) framework for early prediction of customer attrition in e-commerce using machine learning.

Utilized the E-Commerce Customer Churn Dataset and performed comprehensive data preprocessing, including missing value removal, label encoding, Z-score normalization, and SMOTE-based class balancing.

Implemented two advanced ensemble learning models, XGBoost and Random Forest, for accurate customer churn prediction.

Demonstrated that the proposed models outperform conventional machine learning methods such as Logistic Regression, Stochastic Gradient Boosting, Naïve Bayes, Support Vector Machine, and Artificial Neural Network.

Evaluated the proposed models using accuracy, precision, recall, F1-score, and ROC curve to ensure reliable performance assessment.

Provided an effective decision-support framework that enables e-commerce organizations to identify potential churn customers early, improve retention strategies, enhance customer loyalty, and support data-driven CRM decisions.

B. Justification and Novelty

The justification for this study lies in the growing need for accurate, intelligent customer attrition prediction to enhance customer retention and CRM decision-making in e-commerce. The novelty of the proposed work lies in combining complete data preprocessing techniques (Z-score normalization and SMOTE-based class balancing) with advanced ensemble learning models (XGBoost and Random Forest). This approach is a good solution to the problem of class imbalance and enhances predictive power, achieving maximum accuracy. The proposed method is an effective and feasible solution for identifying customers likely to churn early and taking proactive measures to retain them.

C. Organization of the Paper

The remainder of this paper is organized as follows: Section II presents related work on Customer Attrition Prediction in e-commerce, and Section III discusses the methodology and implementation details of the proposed model. Section IV covers the study's results and findings, and Section V discusses the conclusions and future work.

II. LITERATURE REVIEW

A comprehensive review and analysis of key research studies on Intelligent CRM for Customer Attrition Prediction was conducted to inform and enhance the development of this study.

B. T. Giang (2026), proposes a machine learning-based framework for customer churn prediction and retention using demographic, transactional, and behavioral data. Customer segmentation was performed using K-Prototypes and RFM analysis. Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, and SVM were evaluated. Naïve Bayes achieved the highest Recall (94.49%) and F1-score (67.79%), demonstrating its effectiveness for customer churn prediction and targeted retention strategies[32].

Manoharan et al., (2025) competitive market, businesses use artificial intelligence (AI) to deliver personalized recommendations that improve customer engagement, retention, and revenue. Machine learning (ML) models analyze customer behavior and preferences to predict recommendations, product on-time delivery, annual spending, and customer churn. This study investigates the application of ML techniques for these e-commerce predictions and addresses a gap in the literature through empirical analysis. The results demonstrate that ML methods effectively predict customer churn and product delivery, with Gradient Boosting achieving the best performance, reaching 90.67% accuracy for customer churn prediction and 73.75% accuracy for product on-time delivery[33].

Hu and Chen, (2025) proposes a hybrid model combining machine learning and fuzzy logic for efficient customer churn prediction and retention. The model uses a Support Vector Machine (SVM) optimized with the Coyote Optimization Algorithm (COA) for churn prediction, while Fuzzy Z-numbers support customer retention decisions by considering behavioral, demographic, and preference data. The model was evaluated on 8,913 WooCommerce transactions and achieved 97.2% accuracy with 0.98 precision, outperforming previous methods. The proposed approach provides an effective tool for improving customer retention and supporting sustainable growth in e-commerce[34].

Altairey and Al-Alawi, (2024) explores customer churn prediction using four ML algorithms: Logistic Regression, Random Forest, Gradient Boosting, and Histogram Gradient Boosting. A dataset of 7,043 samples with 21 demographic and account-related features was used for training and evaluation. After data preparation and dataset splitting, all four models effectively predicted customer churn. The results also highlight the importance of data balancing. Among the models, Histogram Gradient Boosting achieved the highest accuracy (85.56%), followed by Random Forest (85.41%), Gradient Boosting (84.73%), and Logistic Regression (81.45%)[35].

Faisal et al., (2024) predicts customer churn in dairy products using ML algorithms, including Logistic Regression, Random Forest, Support Vector Machine, K-Nearest Neighbor (KNN), Decision Tree, and Gaussian Naive Bayes. After exploratory data analysis (EDA), churn behavior was analyzed using the Kaplan–Meier curve based on customer lifetime value (CLV). Random Forest and Logistic Regression were selected for detailed evaluation using AUROC. Random Forest achieved the best performance with 87.26% accuracy, 85.44% precision, 88% recall, 86.7% F1-score, and 95% AUC, outperforming Logistic Regression (93% AUC) and proving most effective for dairy product churn prediction[36].

R and N, (2023) examines customer churn in the telecom sector using four classification algorithms: Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), and a Hybrid Algorithm (HA). After preprocessing the telecom dataset to handle missing values and outliers, the models were trained and evaluated. The results showed that the Hybrid Algorithm achieved the highest accuracy (95%), followed by Random Forest (94%), Decision Tree (91%), and

DOI: 10.48175/IJARSCT-37631

Logistic Regression (85%). Performance was assessed using accuracy, precision, and recall, demonstrating that hybrid and Random Forest approaches provide effective customer churn prediction[37].

Galal, Rady and Aref, (2022) proposes a classifier-based model for predicting customer churn in digital banking to support customer relationship management (CRM). Several supervised machine learning algorithms, including K-Nearest Neighbors (KNN), Logistic Regression, AdaBoost, Gradient Boosting, and Random Forest, were evaluated. A unified voting classifier with hyperparameter tuning was used to improve prediction performance, while Random Forest ranked the most important features. Using a 10,000-record dataset, the proposed model achieved 87% accuracy, demonstrating its effectiveness in predicting customer churn[38].

The table I provides an overview of recent research on Intelligent CRM for Customer Attrition Prediction, detailing the proposed models, utilized datasets, major findings, and encountered challenges.

TABLE 1
Recent Studies on Intelligent CRM for Customer Attrition Prediction using Machine Learning technqiues

Author	**Proposed Work**	**Dataset**	**Findings**	**Limitations & Future Work**
B. T. Giang (2026),	Proposed an ML-based churn prediction framework using K-Prototypes clustering, RFM analysis, and multiple classifiers.	Proposed an ML-based churn prediction framework using K-Prototypes clustering, RFM analysis, and multiple classifiers.	Proposed an ML-based churn prediction framework using K-Prototypes clustering, RFM analysis, and multiple classifiers.	Proposed an ML-based churn prediction framework using K-Prototypes clustering, RFM analysis, and multiple classifiers.
Manoharan et al. (2025)	Applied ML models for customer churn, product delivery, and recommendation prediction in e-commerce.	E-commerce customer dataset	Gradient Boosting achieved 90.67% churn prediction accuracy.	Limited evaluation across diverse business domains; future work should investigate deep learning and real-time CRM systems.
Hu and Chen (2025)	Developed a hybrid SVM–COA and Fuzzy Z-number model for churn prediction and retention.	8,913 WooCommerce transactions	Achieved 97.2% accuracy with improved retention decision support.	Requires validation on larger and heterogeneous datasets; future work can explore explainable AI models.
Altairey and Al-Alawi (2024)	Compared Logistic Regression, Random Forest, Gradient Boosting, and Histogram Gradient Boosting.	7,043 customer records	Histogram Gradient Boosting achieved the highest accuracy (85.56%).	Performance depends on data balancing; future research should address class imbalance and feature optimization.
Faisal et al. (2024)	Evaluated multiple ML algorithms for dairy customer churn prediction.	Dairy customer dataset	Random Forest achieved 87.26% accuracy and highest AUC.	Domain-specific dataset limits generalization; future work should test cross-domain applicability.
R and N (2023)	Compared DT, RF, LR, and Hybrid Algorithm for telecom churn	Telecom customer dataset	Hybrid Algorithm achieved 95% accuracy.	Limited algorithm diversity; future studies should evaluate advanced

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, July 2026 Impact Factor: 8.2

	prediction.			ensemble and deep learning techniques.
Galal et al. (2022)	Proposed a voting classifier with hyperparameter tuning for digital banking churn prediction.	10,000 banking customer records	Voting classifier achieved 87% prediction accuracy.	Limited focus on model interpretability and real-time deployment; future work should incorporate explainable and adaptive AI models.

Research gaps: Although previous studies have applied various machine learning techniques for customer attrition prediction, many focus on individual models and report limited predictive performance on diverse e-commerce datasets. The current methods fail to effectively solve the class imbalance issues and fail to incorporate extensive preprocessing steps with advanced ensemble learning techniques. Besides, the use of intelligent CRM with very accurate churn prediction has not been well studied. Therefore, this study proposes a robust framework that combines Z-score normalization, SMOTE, XGBoost, and Random Forest to improve the accuracy and reliability of customer attrition prediction.

III. RESEARCH METHODOLOGY

The methodology proposed starts by acquiring the E-Commerce Customer Churn Dataset then proceeds with data pre-processing, which involves the elimination of missing values, the conversion of categorical data to numerical through label encoding, the normalization of numerical data using Z-score, and the balancing of the classes using SMOTE. The processed dataset is split into training (70%) and testing (30%) sets, and XGBoost and Random Forest models are trained to predict customer attrition. Finally, the models are evaluated using accuracy, precision, recall, F1-score, and ROC curve analysis to identify the most effective approach for intelligent CRM-based customer churn prediction. Fig. 1 illustrates the proposed flowchart for Intelligent CRM for Customer Attrition Prediction using machine learning.

The following section presents a detailed description of each step involved in the proposed methodology:

A. Data Gathering and Analysis

The study uses the E-Commerce Customer Churn Dataset obtained from Kaggle, containing 5,630 customer records with 20 features collected from June 2021 to November 2021. The dataset includes demographic, transaction, shopping behavior, and customer activity attributes for predicting customer churn. Data visualizations such as bar plots and heatmaps were used to examine distribution, feature correlations etc., are given below:

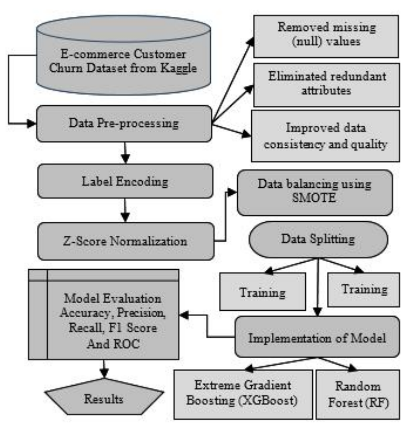

Fig. 1. Proposed flowchart for Intelligent CRM for Customer Attrition Prediction using machine learning

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, July 2026
Impact Factor: 8.2

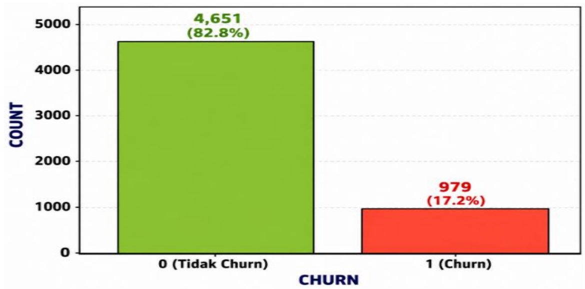

Fig. 2. Bar graph of class distribution

Fig. 2 The bar chart shows that the majority of customers belong to the 0 (Tidak Churn) class, indicating that most customers remained active, while a smaller proportion belongs to the 1 (Churn) class, representing customers who left the e-commerce platform. This class imbalance indicates that there is likely more emphasis on customer retention than on customer attrition in this dataset.

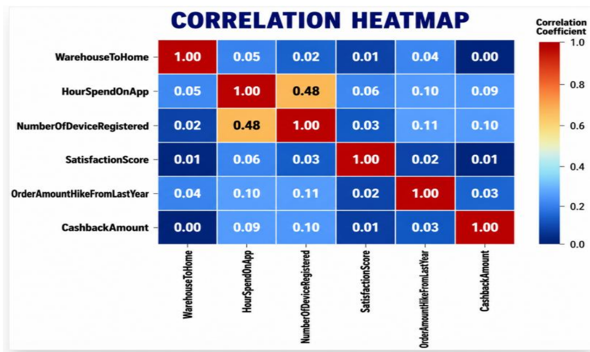

Fig. 3. Correlation Matrix Heatmap

Figure 3 shows the relationships between the selected customer behaviour features, using a colour gradient that runs from red for strong positive correlations, through colour to grey for weak positive and negative correlations, and white for weak negative correlations. It assists in detecting feature relationships and in feature selection by presenting variables with similar correlations prior to machine learning modeling.

B. Data Pre-processing

The E-commerce Customer Churn Dataset was used for data preparation, including cleaning and preprocessing. The preprocessing steps included removing missing (null) values, eliminating redundant attributes, improving data consistency and quality, followed by label encoding and normalization. The main preprocessing procedures are summarized as follows:

- **Removed missing (null) values:** Missing or null values were identified and removed from the dataset. This improved data completeness and prevented errors during model training.
- **Eliminated redundant attributes:** Irrelevant and duplicate attributes were eliminated from the dataset. This reduced unnecessary complexity and enhanced computational efficiency.
- **Improved data consistency and quality:** The dataset was standardized to ensure consistent formatting and reliable records. This enhanced overall data quality and supported accurate machine learning predictions.
- **Label Encoding:** Categorical variables were converted into numerical values using label encoding to make the dataset suitable for machine learning algorithms. This transformation enabled efficient processing of non-numeric features during model training.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, July 2026 Impact Factor: 8.2

C. Z-Score Normalization

Data Normalization is a technique used to transform or standardize the data to have a similar distribution. The most widely used method for data normalization is rescaling or min–max normalization and z-score normalization. In this study, we have applied z-score normalization, a standardization technique with a mean of 0 and having a standard deviation of 1. This scaling technique transforms the values centered around the average value with a unit standard deviation. The z-score normalization is defined as given in Equation (1):

$$E^{'} = \frac{E - \overline{M}}{\sigma_M} \dots \dots \dots (1)$$

Where,
E^' and E are new and old for each data entry, M̅ is the mean, and σ_M is the standard deviation.

D. Data balancing using SMOTE

Data balancing addresses class imbalance by adjusting the class distribution to improve machine learning model performance. In this study, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to generate synthetic samples for the minority (churn) class, reducing model bias and improving the accuracy, recall, and overall predictive performance of the customer churn prediction model.

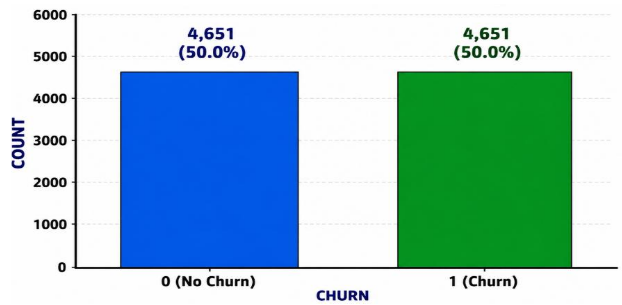

Fig. 4. Bar graph of class distribution SMOTE

Fig. 4 illustrates the balanced class distribution after applying SMOTE, with equal numbers of samples in the No Churn (0) and Churn (1) classes. This balanced dataset reduces class imbalance and enables the machine learning model to predict customer churn more accurately.

E. Data Splitting

The dataset was divided into 70% training data and 30% testing data for model development and evaluation.

F. Implementation of Model

The proposed CRM system for early identification of customer attrition in e-commerce implements Random Forest (RF) and Extreme Gradient Boosting (XGBoost) models on the preprocessed.

A. Proposed Extreme Gradient Boosting (XGBoost) Model

The proposed XGBoost model predicts customer attrition by combining multiple decision trees through a gradient boosting framework. The model constructs trees sequentially, where each new tree minimizes the prediction errors of the previous ensemble. Regularization reduces overfitting and improves generalization. In this study, the XGBoost classifier is trained using the preprocessed customer dataset after feature encoding, normalization, and SMOTE balancing, enabling effective learning of customer attributes. The prediction function of the XGBoost model is expressed as:

DOI: 10.48175/IJARSCT-37631

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \qquad f_k \in F \,\ldots\ldots(1)$$

where y^_irepresents the predicted class of the ith customer, K denotes the total number of decision trees, f_k is the prediction generated by the k^thtree, x_i is the feature vector of the customer, and F represents the space of all regression trees. Equation (1) indicates that the final prediction is obtained by aggregating the outputs of all boosted trees.

To optimize the model while preventing overfitting, XGBoost minimizes the following objective function:

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \ldots\ldots(2)$$

where l(y_i,y^_i)denotes the classification loss between the actual label y_i and the predicted output y^_i, while Ω(f_k) is the regularization term controlling the complexity of each decision tree. Equation (2) ensures that the proposed model achieves high predictive accuracy while maintaining robustness and reducing model complexity.

The trained XGBoost model produces the final customer attrition prediction by integrating the outputs of all optimized trees. Owing to its boosting strategy, regularization mechanism, and efficient handling of non-linear relationships, the proposed model effectively identifies customers at risk of churn and contributes to improved CRM decision-making.

The proposed XGBoost model was configured with n_estimators = 200, max_depth = 6, learning_rate = 0.10, subsample = 0.80, colsample_bytree = 0.80, min_child_weight = 1, gamma = 0, objective = binary, and random_state = 42. These hyperparameter values were selected to improve prediction accuracy while reducing overfitting and enhancing model generalization.

B. Proposed Random Forest Model

The proposed Random Forest (RF) model is used as an ensemble classification technique to predict customer attrition by combining the outputs of multiple decision trees. Each tree is built using bootstrap samples and random feature selection, reducing overfitting and improving prediction stability. The Random Forest classifier is trained on the preprocessed and SMOTE-balanced customer dataset to capture customer behavior patterns. The prediction is obtained through majority voting and is represented as:

$$\hat{y} = mode\{T_1(x), T_2(x), \ldots\ldots, T_N(x)\} \ldots\ldots(3)$$

where y^is the predicted customer class, T_1 (x),T_2 (x),……,T_N (x)represent the predictions of individual decision trees, and N denotes the total number of trees in the forest. Equation (3) shows that the final prediction is determined by the majority vote of all decision trees. Each decision tree is trained using a bootstrap sample of the original dataset, which can be represented as:

$$D_b \subset D, \qquad b = 1,2, \ldots\ldots, N \qquad \ldots\ldots\ldots(4)$$

where D is the original training dataset and D_b represents the bootstrap sample used to construct the b^thdecision tree. Equation (4) enables each tree to learn from different subsets of the data, thereby increasing the diversity and robustness of the ensemble.

The proposed Random Forest model combines the predictions of multiple trees through majority voting to improve classification performance, reduce variance, and enhance model stability for customer attrition prediction. The model was configured with n_estimators = 200, max_depth = 20, min_samples_split = 2, min_samples_leaf = 1, max_features = sqrt, criterion = gini, bootstrap = True, and random_state = 42 to achieve reliable prediction performance.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, July 2026 Impact Factor: 8.2

Evaluation metrics

A confusion matrix is used to evaluate the proposed customer churn prediction model using True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These outcomes are used to compute accuracy, precision, recall, and F1-score, which assess the predictive performance and reliability of the proposed model, as expressed in Equations (5)–(8):

$$Accuracy = \frac{\mathrm{TP + TN}}{\mathrm{TP + Fp + TN + FN}} \ldots..(5)$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP + FP}} \ldots.(6)$$

$$Recall = \frac{\mathrm{TP}}{TP + FN} \ldots\ldots..(7)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \ldots\ldots.(8)$$

Accuracy is the ratio of correctly classified instances to the total number of instances. Precision measures the proportion of true positive predictions among all positive predictions, while Recall represents the proportion of actual positive instances correctly identified by the model. The F1-score is the harmonic mean of precision and recall. The ROC curve evaluates model performance by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at different classification thresholds.

IV. RESULTS AND DISCUSSION

This section presents the experimental setup and performance evaluation of the proposed XGBoost and Random Forest (RF) models for intelligent CRM-based customer attrition prediction. Experiments were conducted using Python 3.10.6 in the Anaconda Jupyter Notebook environment on a system with an Intel Core i9-13900K CPU, 64 GB RAM, NVIDIA RTX 4090 GPU, and Windows 11 Pro. The Kaggle E-commerce Customer Churn Dataset was used for training and testing. The XGBoost model achieved 98.9% accuracy, 96.8% precision, 94.7% recall, and a 95.6% F1-score, outperforming the RF model, which achieved 98.4% accuracy, 97.1% precision, 90.9% recall, and a 94.5% F1-score.

Classifiction results of proposed models for Intelligent CRM for Customer Attrition Prediction

Matrix	XGBoost Model	RF Model
Accuracy	98.9	98.4
Precision	96.8	97.1
Recall	94.7	90.9
F1-score	95.6	94.5

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, July 2026
Impact Factor: 8.2

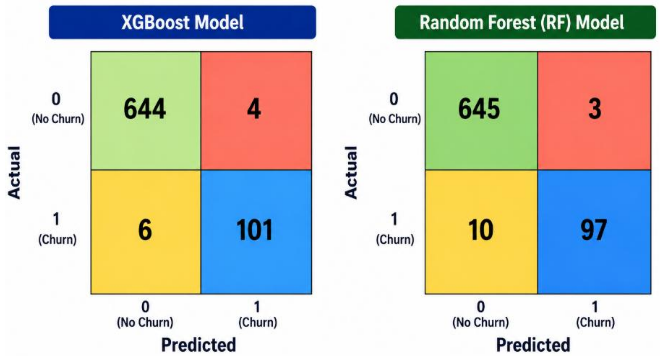

Fig. 5. Confusion Matrix for the Proposed Models

The confusion matrices of the proposed XGBoost and Random Forest (RF) models are presented in Figure 5. Both models correctly classified most No Churn and Churn samples, with only a few misclassifications, demonstrating strong predictive performance and effective customer attrition prediction.

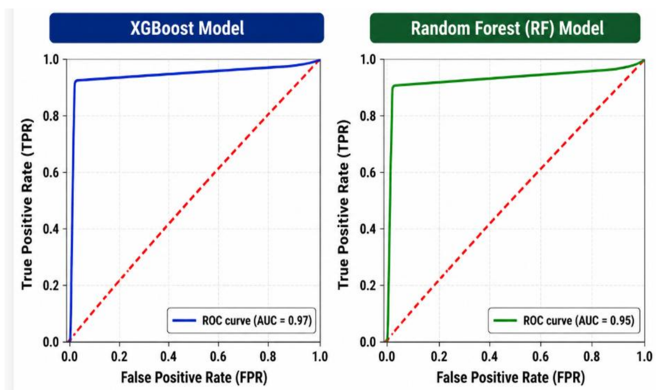

Fig. 6. ROC Curve for the Proposed Models

Figure 6 presents the ROC curves of the proposed XGBoost and Random Forest (RF) models for customer attrition prediction. The XGBoost model achieved an AUC of 0.97, while the RF model obtained an AUC of 0.95, demonstrating excellent classification performance and effective discrimination between churn and non-churn customers.

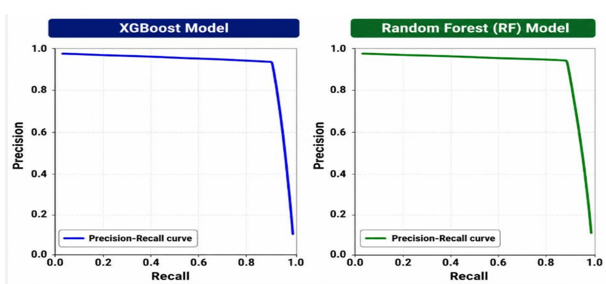

Fig. 8 P-R Curve for Proposed Models

Figure 7 presents the Precision–Recall (P–R) curves of the proposed XGBoost and Random Forest (RF) models. Both models maintain high precision across different recall values, indicating strong classification performance and reliable customer churn prediction with reduced false positive predictions.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, July 2026 Impact Factor: 8.2

A. Comparative analysis

For the effectiveness evaluation of the proposed models a comparative analysis against the other machine learning models is shown in Table III. The comparison is done between Logistic Regression (LR), Stochastic Gradient Boosting (SGB), Naïve Bayes (NB), Support Vector Machine (SVM), and Artificial Neural Network (ANN) employing accuracy, precision, recall and F1-score as evaluation metrics. The proposed XGBoost model achieved an accuracy of 98.9%, precision of 96.8%, recall of 94.7%, and an F1-score of 95.6%, while the proposed Random Forest (RF) model obtained an accuracy of 98.4%, precision of 97.1%, recall of 90.9%, and an F1-score of 94.5%. These results demonstrate that the proposed models provide highly reliable and effective customer attrition prediction, making them well suited for intelligent CRM systems in e-commerce.

Table III. Comparison of Different Machine learning Models for Intelligent CRM for Customer Attrition Prediction

Model	Accuracy	Precision	Recall	F1-score
LR[39]	77	67	41	51
SGB[40]	79	74	69	70
NB[41]	83	83	83	83
SVM[42]	90	64	89	74
ANN[43]	92.09	93	98	95
Proposed XGBoost	98.9	96.8	94.7	95.6
Proposed RF	98.4	97.1	90.9	94.5

The proposed XGBoost and Random Forest (RF) models offer significant advantages for intelligent CRM-based customer attrition prediction by delivering high classification accuracy and reliable customer churn identification. The proposed XGBoost model achieved an accuracy of 98.9%, while the proposed RF model attained 98.4%, demonstrating their ability to effectively learn complex customer behavior patterns and generate dependable predictions. Their strong predictive capabilities help identify customers likely to churn at an early stage, enabling businesses to take proactive steps to retain them, better manage customer relationships, and make informed business decisions.

V. CONCLUSION AND FUTURE STUDY

In recent years, intelligent customer relationship management has become an essential approach for improving customer retention and supporting data-driven decision-making in the rapidly evolving e-commerce industry. This study presented an intelligent CRM framework for the early identification of customer attrition using advanced machine learning techniques. The proposed methodology integrated comprehensive data preprocessing, including label encoding, Z-score normalization, and SMOTE-based class balancing, followed by the implementation of XGBoost and Random Forest classifiers. Experimental results demonstrated that both ensemble models achieved excellent predictive performance, with the proposed XGBoost model outperforming Random Forest by achieving 98.9% accuracy, 96.8% precision, 94.7% recall, 95.6% F1-score, and an AUC of 0.97. These findings confirm that the proposed framework effectively identifies customers at risk of churn, enabling proactive retention strategies, improved customer relationship management, enhanced business decision-making, and increased long-term organizational profitability.

Future research may extend the proposed framework by incorporating deep learning and transformer-based models to further improve prediction performance. Additionally, integrating explainable artificial intelligence (XAI), real-time customer behavioral analytics, and multi-channel data sources can enhance model interpretability, adaptive decision-making, and the effectiveness of intelligent CRM systems in dynamic e-commerce environments.

REFERENCES

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

[2] D. Patel, "Hybrid Deep Learning Approaches for Personalized Product Recommendation in Online Retail Systems," in 2026 6th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), FEZ, Morocco: IEEE, 2026, pp. 1–6, May. doi: 10.1109/IRASET68627.2026.11538894.

[3] A. B. Chatterjee, "A Proposed Hybrid Distributed Ledger Architecture for Cross-Border Payments: Design and Conceptual Framework," Int. J. EmergingTrends Comput. Sci. Inf. Technol., vol. 1, no. 1, pp. 210–217, Feb, 2026, doi: https://doi.org/10.63282/3050-9246.IJETCSIT-V7I1P132.

[4] S. Simaiya et al., "AI driven system for enhancing consumer electronics through maintenance personalization and security," 2026. doi: https://doi.org/10.1038/s41598-026-37401-5.

[5] 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.

[6] T. Shah, "Measuring the Profitability and Long-Term Value of BNPL Customers Using Data Analytics," Int. J. Res. Anal. Rev., vol. 10, no. 1, pp. 505–510, February, 2023.

[7] K. Coussement and K. W. De Bock, "Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning," J. Bus. Res., vol. 66, no. 9, pp. 1629–1636, Sep. 2013, doi: 10.1016/j.jbusres.2012.12.008.

[8] K. Coussement and K. W. De Bock, "Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning," J. Bus. Res., 2013, doi: 10.1016/j.jbusres.2012.12.008.

[9] C. Kudeshia and A. Mittal, "The journey of Chumbak: from mail to mall – leveraging social media networks for on-going dialogue with customers," Emerald Emerg. Mark. Case Stud., vol. 5, no. 4, pp. 1–20, Jul. 2015, doi: 10.1108/EEMCS-09-2014-0209.

[10] S. Singh, A. Mittal, and S. R. George, "Role of Artificial Intelligence in Knowledge Management: An Empirical Study of Industry Experts using Stepwise Multiple Regression," Int. J. Electron. Financ., vol. 12, no. 1, p. 1, 2023, doi: 10.1504/IJEF.2023.10053587.

[11] S. Wu, W.-C. Yau, T.-S. Ong, and S.-C. Chong, "Integrated Churn Prediction and Customer Segmentation Framework for Telco Business," IEEE Access, vol. 9, pp. 62118–62136, 2021, doi: 10.1109/ACCESS.2021.3073776.

[12] D. Zhang, C. Zhang, and C. Zheng, "Prediction and Analysis of Customer Churn of Automobile Dealers Based on BI," in ITNEC 2023 - IEEE 6th Information Technology, Networking, Electronic and Automation Control Conference, 2023. doi: 10.1109/ITNEC56291.2023.10082554.

[13] S. H. D. KOLAGALI, N. S. M. VUPPALAY, and K. S. Hebbar, "Analytical Study of Demographic Bias in LLM-Augmented CRM Systems Using Adaptive Mitigation Layers," Int. J. Soft Comput. Artif. Intell., vol. 13, no. 2, pp. 29–37, Nov, 2025.

[14] B. Jeganathan, "Exploring the Power of Generative Adversarial Networks (GANs) for Image Generation: A Case Study on the MNIST Dataset," Int. J. Adv. Eng. Manag., vol. 7, no. 1, pp. 21–46, Jan. 2025, doi: 10.35629/5252-07012146.

[15] D. Patel, "Integrating Price Elasticity and Reinforcement Learning: A Data-Driven Framework for Strategic E-commerce Pricing," in 2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC), Houston, TX, USA: IEEE, 2026, pp. 1–6, February. doi: 10.1109/ICAIC67076.2026.11395747.

[16] T. P. Patel and P. Agarwal, "Dynamic Multi-Agent Reinforcement Learning for Competitive E-Commerce Price Optimization: Curriculum-Driven SAC, Equilibrium Analysis and Ethical Safeguards," in 2026 International Seminar on Intelligent Business and Edge-Computing Research (ISIBER), 2026, pp. 214-219,February. doi: 10.1109/ISIBER68248.2026.11470206.

[17] 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.

[18] J. Ahn, J. Hwang, D. Kim, H. Choi, and S. Kang, "A Survey on Churn Analysis in Various Business Domains," IEEE Access, vol. 8, pp. 220816–220839, 2020, doi: 10.1109/ACCESS.2020.3042657.

[19] A. Tamaddoni Jahromi, M. M. Sepehri, B. Teimourpour, and S. Choobdar, "Modeling customer churn in a non-contractual setting: the case of telecommunications service providers," J. Strateg. Mark., vol. 18, no. 7, pp. 587–598, Dec. 2010, doi: 10.1080/0965254X.2010.529158.

[20] A. Y. Kusdiyanto and Y. Pristyanto, "Machine Learning Models for Classifying Imbalanced Class Datasets Using Ensemble Learning," in 2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), IEEE, Dec. 2022, pp. 648–653. doi: 10.1109/ISRITI56927.2022.10052887.

[21] 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.

[22] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," IEEE Access, 2019, doi: 10.1109/ACCESS.2019.2914999.

[23] M. Imani, M. Joudaki, A. Beikmohammadi, and H. Arabnia, "Customer Churn Prediction: A Systematic Review of Recent Advances, Trends, and Challenges in Machine Learning and Deep Learning," Mach. Learn. Knowl. Extr., vol. 7, no. 3, p. 105, Sep. 2025, doi: 10.3390/make7030105.

[24] S. Singh, S. A. Pahune, P. Chatterjee, and R. Sura, "Advanced Machine Learning Techniques for Prediction of Customer Churn in Telecommunication Sector," in 2025 IEEE 6th India Council International Subsections Conference (INDISCON), Rourkela, India: IEEE, 2025, pp. 1–6, August. doi: 10.1109/INDISCON66021.2025.11252233.

[25] Y. Patel, "A Reputation-Resilient Seller Ranking Mechanism for Online Marketplaces," Newark J. Human-Centric AI Robot. Interact., vol. 4, pp. 374–393, 2024.

[26] J. W. Sajja and A. Nerella, "Enterprise Finance Reimagined: Harneessing ERP and Data Innovation for Next-Generation Value Creation," 2024. doi: 10.52710/cfs.743.

[27] C. Xie, R. P. Bagozzi, and S. V. Troye, "Trying to prosume: toward a theory of consumers as co-creators of value," J. Acad. Mark. Sci., vol. 36, no. 1, pp. 109–122, Mar. 2008, doi: 10.1007/s11747-007-0060-2.

[28] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," IEEE Access, vol. 7, pp. 60134–60149, 2019, doi: 10.1109/ACCESS.2019.2914999.

[29] A. Amin et al., "Comparing Oversampling Techniques to Handle the Class Imbalance Problem: A Customer Churn Prediction Case Study," IEEE Access, vol. 4, pp. 7940–7957, 2016, doi: 10.1109/ACCESS.2016.2619719.

[30] E. G. Castro and M. S. G. Tsuzuki, "Churn Prediction in Online Games Using Players' Login Records: A Frequency Analysis Approach," IEEE Trans. Comput. Intell. AI Games, vol. 7, no. 3, pp. 255–265, Sep. 2015, doi: 10.1109/TCIAIG.2015.2401979.

[31] S. Charandabi and O. Ghanadiof, "Evaluation of Online Markets Considering Trust and Resilience: A Framework for Predicting Customer Behavior in E-Commerce," J. Bus. Manag. Stud., vol. 4, no. 1, pp. 23–33, Jan. 2022, doi: 10.32996/jbms.2022.4.1.4.

[32] B. T. Giang, N. D. T. Nguyen, T. N. C. Bui, and T. C. N. & D. T. Tran, "Machine learning-based prediction and optimization of customer churn in e-commerce," 2026. doi: https://doi.org/10.1007/s43546-026-01168-8.

[33] G. Manoharan, R. Gupta, B. B, V. V. Darunkar, S. K. Mishra, and M. Soni, "The impact of Artificial Intelligence on E-commerce Customer Experience and Personalization," in 2025 6th International Conference for Emerging Technology (INCET), IEEE, May 2025, pp. 1–5. doi: 10.1109/INCET64471.2025.11140097.

[34] S. Hu and A. Chen, "Data-Driven Customer Retention Strategies in E-Commerce: A Fuzzy Z-Number Approach," IEEE Access, vol. 13, pp. 75388–75399, 2025, doi: 10.1109/ACCESS.2025.3550190.

ISSN: 2581-9429 Volume 6, Issue 1, July 2026 Impact Factor: 8.2

[35] H. A. Altairey and A. I. Al-Alawi, "Customer Churn Prediction in the Telecommunication Industry Using Multiple Machine Learning Algorithms," in 2024 International Conference on Open Innovation and Digital Transformation (OIDT), IEEE, Mar. 2024, pp. 1–10. doi: 10.1109/OIDT59407.2024.11082687.

[36] A. M. Faisal, V. Rajkiran, K. S. Singh, P. Elantheraiyan, and S. K. Kamalakhannan, "Identification of an Efficient Machine Learning Algorithm for the Prediction of Customer Churn: A Case Analysis of a Business-to-Consumer (B2C) Dairy Company for Customer Retention Strategy," in 2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM), 2024, pp. 1–5. doi: 10.1109/ICIPTM59628.2024.10563757.

[37] N. R and L. N, "Machine Learning Methods for Predictive Customer Churn Analysis in the Telecom Industry," in 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), IEEE, Jul. 2023, pp. 1–7. doi: 10.1109/ICCCNT56998.2023.10306395.

[38] M. Galal, S. Rady, and M. Aref, "Enhancing Customer Churn Prediction in Digital Banking using Ensemble Modeling," in NILES 2022 - 4th Novel Intelligent and Leading Emerging Sciences Conference, Proceedings, 2022. doi: 10.1109/NILES56402.2022.9942408.

[39] P. S. Kumar, "Predicting Customer Churn in E-Commerce Using Machine Learning: A Comprehensive Approach," IJIIS Int. J. Informatics Inf. Syst., vol. 9, no. 2, pp. 357–370, Mar. 2026, doi: 10.47738/ijiis.v9i2.300.

[40] B. Prabadevi, R. Shalini, and B. R. Kavitha, "Customer churning analysis using machine learning algorithms," Int. J. Intell. Networks, vol. 4, no. May, pp. 145–154, 2023, doi: 10.1016/j.ijin.2023.05.005.

[41] K. Dhangar and P. Anand, "A Review on Customer Churn Prediction Using Machine Learning Approach," Int. J. Innov. Eng. Res. Technol., vol. 8, no. 05, pp. 193–201, 2021.

[42] D. Amanda Ardhani and K. D. Tania, "Knowledge Discovery on E-Commerce Customer Churn Using Interpretable Machine Learning: A Comparative Study of SHAP-Based Classifiers," J. Appl. Informatics Comput., vol. 9, no. 5, pp. 2695–2702, 2025, doi: 10.30871/jaic.v9i5.10811.

[43] I. Boukrouh and A. Azmani, "Explainable machine learning models applied to predicting customer churn for e-commerce," IAES Int. J. Artif. Intell., vol. 14, no. 1, p. 286, Feb. 2025, doi: 10.11591/ijai.v14.i1.pp286-297