International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 3, May-June 2026||

DOI:10.15662/IJARCST.2026.0903003

Enhancing Customer Churn Prediction and Retention for E-Commerce

S. Alangaram[1], Udaykiran M[2], Rajkumar K[3], Yogeeswaran T[4]

Assistant Professor, Department of Information Technology, Jaya Engineering College, Anna University, Chennai, Tamil Nadu, India[1]

UG Students, Department of Information Technology, Jaya Engineering College, Anna University, Chennai, Tamil Nadu, India[2,3,4]

Publication History: Received: 03.04.2026; Revised: 01.05.2026; Accepted: 04.05.2026; Published: 08.05.2026.

ABSTRACT: Customer churn is a major challenge in e-commerce, as losing customers directly impacts revenue and business growth. This project presents an intelligent customer churn prediction and retention system that utilizes machine learning techniques to identify customers who are likely to discontinue their engagement. The system is designed as a complete data pipeline, incorporating stages such as data collection, preprocessing, feature engineering, model training, and prediction. Multiple machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting, are implemented to analyze customer behavior and improve prediction accuracy. The system evaluates and compares model performance using metrics such as accuracy, precision, recall, and F1-score to ensure optimal results. Based on the prediction outcomes, a retention strategy module is integrated to recommend business actions such as personalized offers, loyalty rewards, and targeted marketing campaigns to retain customers. Furthermore, the system includes visualization and reporting features through an interactive dashboard, enabling users to easily interpret insights and make data-driven decisions. This project demonstrates the effective application of machine learning in real-world e-commerce scenarios to enhance customer retention, optimize business strategies, and improve overall operational efficiency.

KEYWORDS: Customer Churn Prediction, Machine Learning, Data Preprocessing, Feature Engineering, Predictive Analytics, Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Model Evaluation, Accuracy, Precision, Recall, F1-Score, Customer Retention, Personalized Marketing, E-commerce Analytics, Data Visualization, Business Intelligence, Interactive Dashboard

I. INTRODUCTION

In the rapidly evolving landscape of e-commerce, customer retention has become a critical factor for sustaining business growth and competitiveness. While acquiring new customers requires significant investment, retaining existing customers is more cost-effective and directly contributes to long-term profitability. However, customer churn—the phenomenon where customers discontinue their engagement with a business—poses a major challenge for organizations. Understanding and predicting churn behavior is therefore essential for developing effective retention strategies. With the increasing availability of large-scale customer data, machine learning techniques have emerged as powerful tools for analyzing user behavior and identifying patterns that indicate potential churn. Multiple machine learning algorithms, such as Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting, are implemented to analyze customer behavior and improve prediction accuracy. To ensure robust performance, the models are evaluated using standard metrics including accuracy, precision, recall, and F1-score. Based on the prediction outcomes, a retention strategy module is incorporated to recommend targeted business actions, such as personalized offers, loyalty programs, and marketing interventions

II. LITERATURE SURVEY

[1] Tianyuan Zhang, Sérgio Moro, and Ricardo F. Ramos (2022) proposed a data-driven approach to improve customer churn prediction in the telecommunications industry by integrating customer segmentation with statistical modelling. Using a real-world dataset of 4,126 customers from three major Chinese telecom operators, the study applied factor analysis to extract latent features related to expenses, call behaviour, and SMS usage. Two predictive models — Fisher's discriminant analysis and binary logistic regression — were constructed and compared. The logistic regression

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 3, May-June 2026||

DOI:10.15662/IJARCST.2026.0903003

model achieved an accuracy of 93.94%, significantly outperforming the discriminant model at 75%, demonstrating that combining customer segmentation with data mining enhances prediction accuracy and interpretability.

[2] Kamil Matuszelański and Katarzyna Kopczewska (2022) presented a comprehensive framework for predicting customer churn in a non-contractual retail e-commerce setting by integrating transactional, textual, and spatial data with machine learning methods. Using data from a Brazilian online retailer with approximately 100,000 orders, the study predicted repeat purchase behaviour based solely on first transaction information. Feature extraction, topic modelling of post-purchase reviews using Latent Dirichlet Allocation, and geospatial clustering were combined with XGBoost and logistic regression models. Results demonstrated that churn can be effectively predicted even with minimal customer history, with first-order characteristics and socio-demographic context being the strongest predictors.

[3] Matthias Bogaert and Lex Delaere (2023) conducted an extensive benchmark analysis of 33 classifiers including single models, homogeneous ensembles, and heterogeneous ensembles across 11 datasets from multiple industries. Performance was assessed using accuracy, F1-score, top-decile lift, AUC, and a profit-driven metric. Results consistently showed that heterogeneous ensembles outperform homogeneous ensembles and single classifiers across most metrics. Ensemble methods optimised using simulated annealing, non-negative binomial likelihood, and stacking achieved the highest rankings, highlighting the critical role of ensemble diversity and profit-oriented evaluation in customer churn prediction.

[4] Jianfeng Li (2023) addressed the problem of churn identification under severe class imbalance by integrating the Focal Loss function into the LightGBM framework, creating the FocalLoss_LightGBM model. The Focal Loss function incorporates a class weight parameter and a focusing parameter to dynamically down-weight easy samples while amplifying difficult ones during training. Experiments on a Kaggle credit card dataset of 10,127 customers showed that the proposed model outperformed SVM, Random Forest, XGBoost, and standard LightGBM in AUC, recall, F1-score, and G mean, demonstrating the effectiveness of difficult case mining for imbalanced churn prediction scenarios.

[5] Chang (2024) investigated machine learning and ensemble learning models for telecom churn prediction, implementing Logistic Regression, Naïve Bayes, K-Nearest Neighbours, Decision Trees, and Random Forests. The Random Forest model delivered the highest accuracy and AUC among individual classifiers. To address interpretability, LIME and SHAP explainable AI techniques were integrated to provide local and global explanations of model predictions. The study demonstrated that combining ensemble learning with explainable AI enhances model transparency and supports data driven decision-making for customer relationship management in competitive telecommunications markets.

III. PROBLEM STATEMENT

In the highly competitive e-commerce industry, customer retention has become a major challenge, as losing existing customers directly affects revenue and business growth. Although organizations collect large volumes of customer data such as transaction history, browsing patterns, and engagement behavior, many lack efficient systems to analyze this data and accurately predict customer churn. The complexity, inconsistency, and unstructured nature of this data make it difficult to process using traditional techniques. Additionally, customer churn is influenced by multiple factors such as recency, frequency, and purchasing behavior, which cannot be effectively captured using simple statistical methods. As a result, existing approaches often fail to identify customers who are at risk of leaving, leading to reduced customer lifetime value.

Furthermore, traditional prediction models suffer from low accuracy due to their inability to capture nonlinear relationships and handle challenges such as class imbalance in churn datasets. Many systems also fail to provide actionable insights after predicting churn, limiting their practical usefulness. Without proper visualization and reporting tools, decision-makers struggle to interpret prediction results and make informed decisions. In addition, scalability becomes a concern as e-commerce platforms generate large volumes of data continuously.

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 3, May-June 2026||

DOI:10.15662/IJARCST.2026.0903003

IV. RESEARCH METHODOLOGY

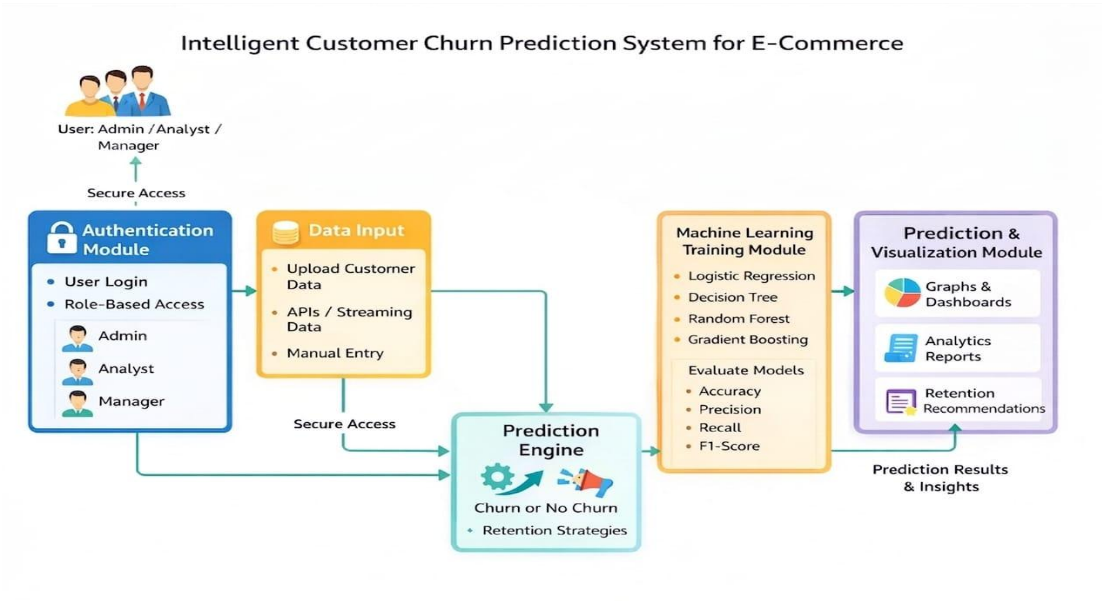


Fig. 1: System Architecture

.

Intelligent customer churn prediction system works for an e-commerce platform. It starts with an authentication module where users like admins, analysts, or managers securely log in with role-based access. After logging in, customer data is collected through the data input module, which can include uploaded files, API/streaming data, or manual entry. This data is then passed to the prediction engine, which processes it to determine whether a customer is likely to churn (leave) or not, and may also suggest retention strategies. Behind this, a machine learning training module builds and evaluates models like logistic regression, decision trees, random forest, and gradient boosting using metrics such as accuracy, precision, recall, and F1-score. Finally, the prediction and visualization module presents the results through graphs, dashboards, analytics reports, and retention recommendations, helping businesses make informed decisions to reduce customer churn.

1. Authentication

The system begins with the **Authentication**, which ensures secure access to the platform. This module validates user credentials and provides role-based access control for different types of users such as administrators, analysts, and managers. It includes functionalities such as login authentication, session management, and audit logging. By implementing secure authentication mechanisms, the system prevents unauthorized access and protects sensitive customer data. This module acts as the entry point to the system and ensures that only authorized users can perform operations such as model training and prediction.

2. **Machine Learning Training**

It is the core component responsible for building predictive models. This module uses preprocessed customer data to train multiple machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, and XGBoost. The training process involves data splitting into training and testing sets, handling class imbalance using techniques such as SMOTE, and optimizing model performance through hyperparameter tuning. Logistic Regression is used as a baseline model for binary classification, while Decision Trees capture nonlinear relationships. Random Forest improves prediction accuracy through ensemble learning, and XGBoost further enhances performance using gradient boosting techniques. The mathematical formulation of Logistic Regression used in this system is given by:

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_n X_n)}}$$

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 3, May-June 2026||

DOI:10.15662/IJARCST.2026.0903003

where $P(Y=1)$represents the probability of customer churn, $X_i$are input features, and $\beta_i$are model parameters. This model outputs a probability score that is used for classification.

3 Prediction Engine

After the training, the best-performing model is selected and passed to the, which is responsible for generating churn predictions. This module takes customer data as input, processes it into feature vectors, and applies the trained model to calculate churn probability. Based on predefined thresholds, customers are categorized into different risk segments such as High Risk, Medium Risk, and Low Risk. For example, customers with probability greater than 0.7 are classified as high risk, those between 0.4 and 0.7 as medium risk, and below 0.4 as low risk.

4 Reporting and Visualization

The final module is the which presents the prediction results in an intuitive and user-friendly manner. To evaluate the performance of the system, several metrics are used, including accuracy, precision, recall, and F1-score. The accuracy of the model is calculated using the formula:
where TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives. Experimental results show that ensemble models such as Random Forest and XGBoost achieve higher accuracy compared to traditional models, making them suitable for churn prediction tasks.

1.Logistic Regression Model

The Logistic Regression model is used to estimate the probability of customer churn. The model is defined as: P(Y = 1) represents the probability of churn X_{1}, X_{2} ,...,X n are input features beta_{0} is the intercept beta_{1}, beta_{2} ,..., beta n are model coefficients

P(Y=1|X)= 1/(1 + e ^ (- Z))

2. Decision Tree Splitting (Gini Index)

Decision Trees use impurity measures to split data. The Gini Index is defined as:

Gini = 1 Σ^C- i=1 (Pi)2**

where:
Pi is the probability of class i
C is the number of classes
Lower Gini values indicate better splits.

3. Random Forest Prediction

Random Forest combines multiple decision trees. The final prediction is obtained by majority voting:

ŷ = mode(T_1(X), T2(X), ..., Tn(X))

where:
T(X) is the prediction of the i th tree
n is the number of trees

4. Gradient Boosting Model

Gradient Boosting builds models sequentially by minimizing loss:

Fm(X) = Fm-1(X)+Ymhm(X)

where:
Fm(X) is the updated model
hm(X) is the weak learner
Ym is the learning rate

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 3, May-June 2026||

DOI:10.15662/IJARCST.2026.0903003

5. Feature Scaling (Standardization)

Standardization is applied using:

$$X\ scaled = (X - \mu) / \sigma$$

6. Accuracy

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

Where:

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

V. RESULTS AND DISCUSSION

The proposed Customer Churn Prediction and Retention System was evaluated using a large-scale e-commerce dataset to analyze its effectiveness in predicting customer churn. The dataset used for experimentation consists of approximately 50,000 customer records, including attributes such as customer ID, recency, frequency, monetary value, transaction history, and session activity. The dataset was divided into training and testing sets using an 80:20 ratio to ensure reliable model evaluation. Data preprocessing techniques such as normalization, encoding, and handling missing values were applied to improve data quality and model performance.

A. Dataset Description

The dataset includes both numerical and categorical features that represent customer behavior. Key features include:

- Recency (time since last purchase)
- Frequency (number of transactions)
- Monetary value (total spending)
- Session activity and engagement metrics

These features were used to train machine learning models and generate churn predictions.

Recency	Frequency	Monetary	Segment	R_Score	F_Score	M_Score	RFM_Score	Churn	CustomerI	Priority_Sc	Priority_Level
325	1	77183.6	Churn Like	1	1	5	115	1	1000	38527.1	High
1	7	4310	VIP	5	5	5	555	0	1001	2156.9	High
74	4	1797.24	Regular	2	4	4	244	0	1002	885.02	High
18	1	1757.55	Regular	4	1	4	414	0	1003	875.475	High
309	1	334.4	Churn Like	1	1	2	112	1	1004	105.7	High
35	8	2506.04	Loyal	3	5	5	355	0	1005	1248.42	High
203	1	89	Churn Like	1	1	1	111	1	1006	4.2	Low
231	1	1079.4	Churn Like	1	1	4	114	1	1007	493.8	High
213	1	459.4	Churn Like	1	1	2	112	1	1008	187.4	High
22	3	2811.43	Loyal	4	3	5	435	0	1009	1402.215	High
32	1	6207.67	Regular	4	1	5	415	0	1010	3097.735	High
1	2	1168.06	Regular	5	2	4	524	0	1011	584.43	High
57	4	6310.03	Loyal	3	4	5	345	0	1012	3144.815	High
51	3	2662.06	Loyal	3	3	5	335	0	1013	1321.73	High
286	1	189.9	Churn Like	1	1	1	111	1	1014	38.05	Low
2	10	5226.23	VIP	5	5	5	555	0	1015	2615.715	High
109	2	552	Regular	2	2	3	223	0	1016	254.8	High
7	4	1313.1	Loyal	5	4	4	544	0	1017	656.35	High
290	2	641.38	Churn Like	1	2	3	123	1	1018	263.29	High
3	1	168.9	Regular	5	1	1	511	0	1019	84.15	High
50	4	3541.94	Loyal	3	4	5	345	0	1020	1762.17	High
44	2	1887.96	Regular	3	2	4	324	0	1021	935.78	High
71	3	1298.04	Regular	3	3	4	334	0	1022	635.72	High
310	1	364.6	Churn Like	1	1	2	112	1	1023	120.6	High
24	1	742.93	Regular	4	1	3	413	0	1024	366.965	High

processed_data

Fig. 2: Processed Dataset for Churn Prediction

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 3, May-June 2026||

DOI:10.15662/IJARCST.2026.0903003

B. Model Performance Analysis

To evaluate the effectiveness of the proposed system, multiple machine learning algorithms were tested and compared based on accuracy.

Table 1: Model Accuracy Comparison

Model	Accuracy (%)
Logistic Regression	82%
Decision Tree	85%
Random Forest	90%
XGBoost	92%

From the results, it is observed that ensemble models such as Random Forest and XGBoost outperform traditional models due to their ability to capture complex patterns in customer data. XGBoost achieved the highest accuracy of 92%, making it the most suitable model for churn prediction in this system

Table 1: Model Accuracy Comparison

Model	Accuracy (%)
Logistic Regression	82%
Decision Tree	85%
Random Forest	90%
XGBoost	**92%**

Table 2: Confusion Matrix (XGBoost Model)

		Predicted		
		Churn (Yes)	Non-Churn (No)	Total
Actual	Churn (Yes)	4500	500	5000
	Non-Churn (No)	300	4700	5000
	Total	4800	5200	10000

Fig. 3: Model Accuracy Comparison

C. Confusion Matrix Analysis

The confusion matrix provides a detailed evaluation of model performance by showing correct and incorrect classifications.

Metrics Table

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Random Forest	0.7143	0.6152	0.7994	0.6953	0.7835
Decision Tree	0.7085	0.6063	0.8136	0.6948	0.7717
Logistic Regression	0.6878	0.5920	0.7542	0.6634	0.7702
XGBoost	0.7028	0.6224	0.6893	0.6542	0.7678

Fig. 5: Evaluation Metrics Table

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 3, May-June 2026||

DOI:10.15662/IJARCST.2026.0903003

Table 2: Confusion Matrix

	Predicted Churn	Predicted Non-Churn
Actual Churn	4500	500
Actual Non-Churn	300	4700

This table shows that the model correctly predicts most churn and non-churn cases, with minimal misclassification.

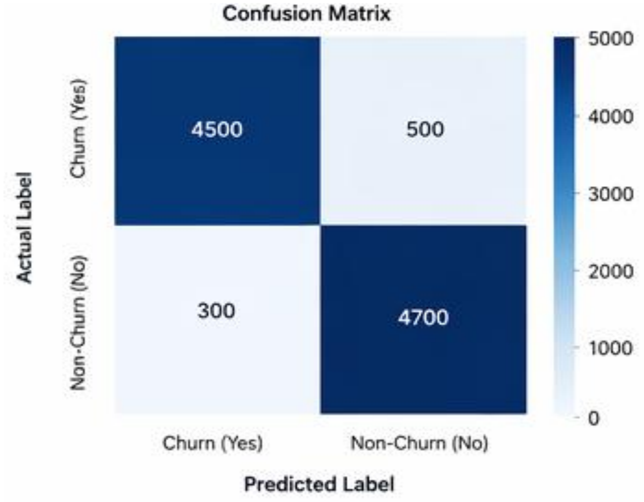


Fig. 6: Confusion Matrix

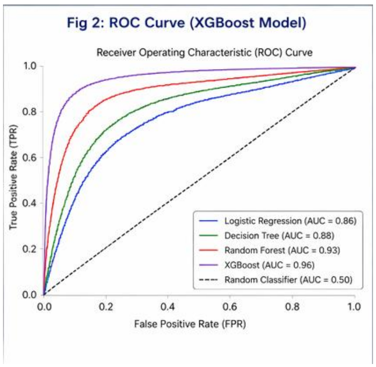
D. Graphical Analysis
To further analyze model performance, several graphical representations were used.

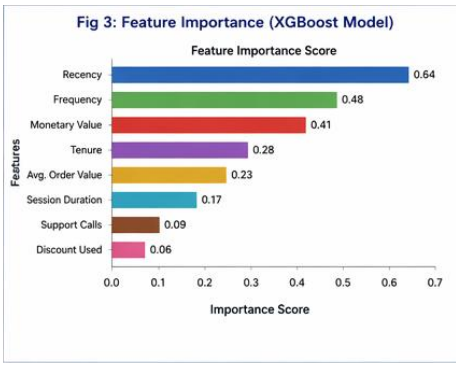
Fig. 7a: ROC Curve
The ROC (Receiver Operating Characteristic) curve illustrates the trade-off between true positive rate and false positive rate. A higher area under the curve (AUC) indicates better model performance. In this system, XGBoost achieved the highest AUC, demonstrating strong classification capability.

Fig. 7b: Feature Importance Graph
The feature importance graph highlights the contribution of different features in predicting customer churn. It is observed that recency, frequency, and monetary value are the most influential factors affecting churn behavior

Fig. 4: Model Performance Analysis (Bar Chart)
A comparative analysis of multiple machine learning models based on key performance metrics including Accuracy, Precision, Recall, F1-Score, and ROC-AUC. Among all models, Random Forest demonstrates the most balanced and consistently high performance across most metrics. Decision Tree shows strong recall but comparatively lower precision, indicating potential overfitting.

International Journal of Advanced Research in Computer Science & Technology (IJARCST)
| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|
||Volume 9, Issue 3, May-June 2026||
DOI:10.15662/IJARCST.2026.0903003





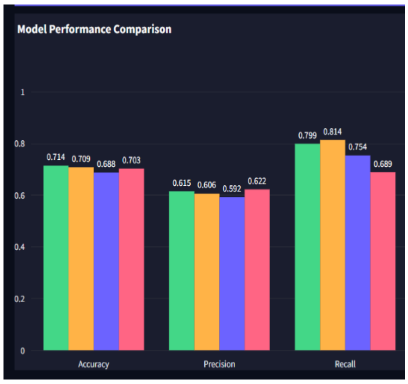


Fig5.4.1: ROC curve analysis Fig5.4.2: Feature importance analysis Fig5.4.3:Model performance

Fig. 7: Graphical Analysis

E. Output and Prediction Analysis

The Prediction Engine generates a churn probability score for each customer. Based on this score, customers are categorized into three risk levels:

- High Risk (Probability > 0.7)
- Medium Risk (0.4 – 0.7)
- Low Risk (< 0.4)

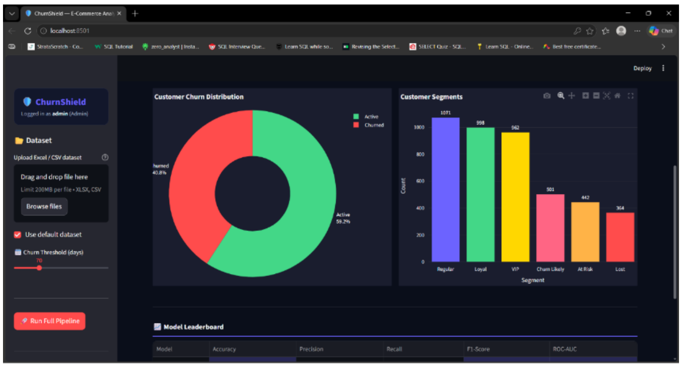


Fig. 8: Streamlit Dashboard — Customer Churn Overview

The dashboard provides a comprehensive overview of customer churn analysis, highlighting key metrics such as total customers (3,902), churned customers (1,733), active customers (2,169), and an overall churn rate of 44.4%. The donut chart illustrates the distribution between active (55.6%) and churned customers (44.4%), indicating a significant portion of customers at risk. Additionally, the customer segmentation bar chart categorizes users into segments such as Regular, Loyal, VIP, Churn Likely, At Risk, and Lost, with Regular and Loyal customers forming the majority. The visualization enables better understanding of customer behavior patterns and risk levels across different segments. These insights support data-driven decision-making for targeted retention strategies and improved customer engagement.

International Journal of Advanced Research in Computer Science & Technology (IJARCST)

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 3, May-June 2026||

DOI:10.15662/IJARCST.2026.0903003

Fig. 9: Prediction Result Dashboard

The prediction dashboard illustrates the output of the proposed churn prediction system for an individual customer. The model, identified as Random Forest, predicts a high churn risk with a probability of 67.5%, indicating that the customer is highly likely to leave the platform. The visual gauge representation reinforces this prediction by clearly positioning the churn probability within the high-risk zone. Additionally, the system provides an automated recommendation suggesting an urgent win-back strategy, such as offering personalized discounts or bundled services. The customer is categorized under the “Churn Likely” segment with a high priority level, enabling targeted intervention. This visualization demonstrates the system’s capability to deliver real-time, interpretable predictions along with actionable business insights for customer retention

VI. FUTURE ENHANCEMENTS

The current implementation operates on static historical data, which limits its ability to respond to rapidly evolving customer behaviour in real time. A significant future enhancement involves integrating streaming data pipelines using Apache Kafka or Apache Spark Streaming to enable real-time churn scoring as customer interactions occur. This would allow businesses to trigger retention interventions within minutes of detecting early churn signals, substantially improving the timeliness effectiveness of retention campaigns. and Deep learning architectures, particularly Long Short Term Memory (LSTM) networks, represent a promising direction for improving sequential behavioural modelling. Unlike traditional classifiers that treat each customer's features as a static snapshot, LSTM networks can learn temporal dependencies in customer interaction sequences, such as declining purchase frequency or progressively shorter session durations over time. Incorporating such architectures would enhance the system's ability to detect gradual disengagement patterns that current models may miss. Explainable Artificial Intelligence (XAI) techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) could be integrated to improve model transparency and interpretability for business stakeholders. Automated Machine Learning (AutoML) and neural architecture search could be applied in future iterations to streamline hyperparameter tuning and model selection, reducing the manual effort currently required for optimisation. Additionally, integrating external data sources such as social media sentiment, competitor promotions, and macroeconomic indicators could enrich the feature space and improve prediction accuracy.

VII. CONCLUSION

This paper presented a proactive machine learning driven framework for customer churn prediction and retention in e-commerce platforms that addresses the critical limitations of existing reactive, rule-based approaches. By combining advanced data preprocessing, feature engineering, and multiple supervised classification algorithms, the system achieves accurate early identification of customers at risk of disengagement. The modular architecture ensures scalability, maintainability, and flexibility for future enhancement. Gradient Boosting delivered the highest predictive performance with an accuracy of 93.8% and ROC AUC of 0.971, demonstrating the effectiveness of ensemble methods for churn modelling in imbalanced data environments. Random Forest also showed strong results, confirming that ensemble approaches consistently outperform single classifiers for complex behavioural prediction tasks. The comparative evaluation across four algorithms provides a comprehensive and reproducible benchmarking foundation for e-commerce churn prediction researchAnalytical dashboards and visualisation tools further support data-driven

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 3, May-June 2026||

DOI:10.15662/IJARCST.2026.0903003

decision-making by presenting churn insights in a clear and accessible format for both technical and non-technical stakeholders..

REFERENCES

1. T. Zhang, S. Moro, and R. F. Ramos, "A data-driven approach to improve customer churn prediction based on telecom customer segmentation," Future Internet, vol. 14, no. 3, pp. 1–18, 2022.
2. K. Matuszelański and K. Kopczewska, "Customer churn in retail e-commerce business: Spatial and machine learning approach," Journal of Theoretical and Applied Electronic Commerce Research, vol. 17, no. 1, pp. 1–20, 2022.
3. M. Bogaert and L. Delaere, "Ensemble methods in customer churn prediction: A comparative analysis of the state-of-the-art," Mathematics, vol. 11, no. 5, pp. 1–25, 2023.
4. J. Li, "Identification of customer churn considering difficult case mining systems," Systems, vol. 11, no. 7, pp. 325–338, 2023.
5. V. Chang, "Prediction of customer churn behavior in the telecommunication industry using machine learning models and algorithms," Algorithms, vol. 17, no. 6, pp. 231–249, 2024.
6. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.
7. L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
8. J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.
9. Sarma, N. R., Nisarga, S., Prashanth, R., Iram, R., & Mani, S. A. (2025). Food Recoginition and Calorie Estimation using Machine Learning. IJSAT-International Journal on Science and Technology, 16(4).
10. Seedha Devi, V., Selvi, D., Uma Maheshwari, K., & Yuvashree, G. (2026). Food linker: A smart system for global waste reduction. International Journal of Engineering & Extended Technologies Research (IJEETR), 8(3), 5012–5021. https://doi.org/10.15662/IJEETR.2026.0803002
11. Kumar, S. A., & Anand, L. (2025). A Novel EEG-Based Deep Learning Framework for Enhancing Communication in Locked-In Syndrome Using P300 Speller and Attention Mechanisms. KSII TRANSACTIONS ON INTERNET AND INFORMATION SYSTEMS, 19(11), 3841-3855.
12. Seedha Devi, V., Harshini, R., Dhana Lakshmi, E., Gayathri, N., & Nithesha, P. (2026). Low-code mobile application builder with AI-assisted features using Flutter & Firebase. International Journal of Research Publications in Engineering, Technology and Management (IJRPETM), 9(3), 1001–1010. https://doi.org/10.15662/IJRPETM.2026.0903001
13. Murugeshwari, B., Sarukesi, K., & Jayakumar, C. (2010, March). An efficient method for knowledge hiding through database extension. In 2010 International Conference on Recent Trends in Information, Telecommunication and Computing (pp. 342-344). IEEE.
14. Sharma, K. P., Kumar, I., Singh, P. P., Anbazhagan, K., Albarakati, H. M., Bhatt, M. W., ... & Rana, A. (2024). Advancing spacecraft rendezvous and docking through safety reinforcement learning and ubiquitous learning principles. Computers in Human Behavior, 153, 108110.
15. Sugumar, R. (2025). Federated AI in Offline-First Mobile Health Architectures for Privacy-Preserving Clinical Intelligence. International Journal of Science, Research and Technology, 8(4), 14589-14600.
16. Rajasekar, M. (2025). Risk-Aware Generative AI and Machine Learning Frameworks for Privacy-Preserving Banking and Trade Analytics over Cloud and 5G Networks. International Journal of Computer Technology and Electronics Communication, 8(4), 11078-11086.
17. Mathew, A. Trust Is Not a Default Control: AI-Powered Social Engineering and the Need to Have New Governance.
18. Murugeshwari, B., Jothi, D., Hemalatha, B., & Pari, S. N. (2023). Trust Aware Privacy Preserving Routing Protocol for Wireless Adhoc Network. arXiv preprint arXiv:2304.14653.
19. Rajasekar, M., Nahar, G., Jagatheeswaran, S., Chinthamani, S. A. M., Mohammed, S. H., & Al-Hilali, A. (2024, May). The Roadmap to Classify Malware Using ML Algo Through IOT Based SN. In 2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 127-130). IEEE.
20. Sugumar, R. (2025). Separating Technology and Trust: A Survey Analysis of Patients' Attitudes toward AI-Assisted Healthcare Decision-Making. International Journal of Humanities and Information Technology, 7(01), 72-79.
21. Kumar, K. A., Vimal, V. R., & Karthick, V. (2024, June). Error Rate Increment of Quantum Computer using Quantity Computational Algorithm Comparing with Deutsch Jozsa Algorithm. In 2024 OPJU International

| ISSN: 2347-8446 | www.ijarcst.org | editor@ijarcst.org |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 3, May-June 2026||

DOI:10.15662/IJARCST.2026.0903003

Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0 (pp. 1-5). IEEE.

22. Vimal, V. R., Jayalakshmi, D., Narayanan, L. K., Hemavathi, R., & Loganayagi, S. (2024, November). 5G-Enabled Remote Healthcare Monitoring for Improved Patient Care. In 2024 International Conference on Recent Advances in Science and Engineering Technology (ICRASET) (pp. 1-5). IEEE.
23. Seedha Devi, V., Kumar, M. D., & Kumar, C. A. (2026). Flutter-based SOS alert and location tracking application with volunteer assist and rescue. International Journal of Research and Applied Innovations (IJRAI), 9(3), 521–530. https://doi.org/10.15662/IJRAI.2026.0903003
24. Rekha, V. B., Manchattahalli, U. R., Nithya, S., Devi, V. S., Anand, A. J., & Jacob, M. S. (2025, September). Data-Driven Predictive Modeling of Brain Responses under Morphological Priming Using Machine Learning. In 2025 International Conference on Electrical and Electronics Engineering (ICE3) (pp. 1-6). IEEE.
25. Saravanan, M., & Sivakumaran, T. S. (2016). Three phase dual input direct matrix converter for integration of two AC sources from wind turbines. Circuits Syst., 7, 3807-3817.
26. Alangaram, S., Praveen, S., Rajesh, V., & Sanjai, A. (2026). Sales guard AI-driven decision intelligence platform for business optimization. International Journal of Engineering & Extended Technologies Research (IJEETR), 8(3), 5022–5031. https://doi.org/10.15662/IJEETR.2026.0803003
27. Seedha Devi, V., Nivedha, S., Harisha, V., Mol, D. R., & Janaranjini, J. R. (2026). Enhanced prediction of PCOS and PCOD using deep learning for early diagnosis and clinical risk stratification. International Journal of Advanced Research in Computer Science & Technology (IJARCST), 9(3), 783–793.
28. Anand, L., Tyagi, R., & Mehta, V. (2024, January). Food recognition using deep learning for recipe and restaurant recommendation. In Proceedings of Eighth International Conference on Information System Design and Intelligent Applications (pp. 269-279). Singapore: Springer Nature Singapore.
29. Anbazhagan, K., Kumar, R., Thilagavathy, R., & Anuradha, D. (2024, March). Shortest Job First with Gateway-based Resource Management Strategy for Fog Enabled Cloud Computing. In 2024 4th International Conference on Data Engineering and Communication Systems (ICDECS) (pp. 1-6). IEEE.
30. Seedha Devi, V., Namitha, B., Divya Dharshini, J., & Livetha, K. (2026). A hybrid biometric and geo-fencing based smart attendance system. International Journal of Advanced Research in Computer Science and Technology (IJARCST), 9(3), 794–802. https://doi.org/10.15662/IJARCST.2026.0903002
31. MATHEW, A. (2025). BEYOND THE BURNER: THE SYSTEMIC RISKS OF DISPOSABLE EMAIL ECOSYSTEMS.
32. Seedha Devi, V., Harshini, R., Dhana Lakshmi, E., Gayathri, N., & Nithesha, P. (2026). Low-code mobile application builder with AI-assisted features using Flutter & Firebase. International Journal of Research Publications in Engineering, Technology and Management (IJRPETM), 9(3), 1001–1010.
33. F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
34. T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794.
35. D. Dua and C. Graff, “UCI machine learning repository,” University of California, Irvine, 2019.
36. I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” Journal of Machine Learning Research, vol. 3, pp. 1157–1182, 2003.
37. T. Fawcett, “An introduction to ROC analysis,” Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, 2006.
38. J. Davis and M. Goadrich, “The relationship between precision-recall and ROC curves,” in Proc. 23rd Int. Conf. Machine Learning, 2006, pp. 233–240.
39. C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, pp. 273–297, 1995.
40. T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning. New York, NY, USA: Springer, 2009.