International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 11 | Nov 2026 www.irjet.net p-ISSN: 2395-0072

ChurnGuard AI: An Explainable Machine-Learning System for Customer Churn Prediction and Retention Support

Shaikh zaid[1], Yash puse[2], Gaurav shedge[3], PROF. J. S. Jadhav[4]

[1,2,3] (Students, Department of Computer Engineering), International Centre of Excellence in Engineering & Management
[4](Professor, Dept. of Computer Engineering), International Centre Of Excellence in Engineering & Management

---

Abstract - *Customer churn is a major challenge for subscription-based businesses because losing customers can reduce recurring revenue and increase customer acquisition costs. This paper presents ChurnGuard AI, a machine-learning-based customer churn prediction and retention-support application developed using Python and Streamlit. The system uses the IBM Telco Customer Churn sample dataset, containing 7,043 customer records, to train and evaluate Logistic Regression, Random Forest, and Extreme Gradient Boosting (XGBoost) classifiers. The preprocessing workflow includes missing-value handling, categorical encoding, and stratified train-test splitting. The models are evaluated using accuracy, precision, recall, F1-score, and ROC-AUC. In the latest reported training run, XGBoost achieved 80.06% accuracy, 65.55% precision, 52.41% recall, 58.25% F1-score, and 84.43% ROC-AUC. The application supports individual customer prediction, bulk CSV scoring, model comparison, business insights, SHAP-based explanations, rule-based retention recommendations, and SQLite-based prediction history. Although XGBoost achieved the highest reported accuracy and ROC-AUC, its recall indicates that further threshold optimization and validation are necessary. ChurnGuard AI is designed as a decision-support application and does not automatically contact customers or guarantee churn prevention.*

Keywords: Customer Churn, Machine Learning, XGBoost, Logistic Regression, Random Forest, SHAP, Streamlit, Explainable AI, Retention Analytics.

1. INTRODUCTION

Customer churn refers to the reduction in a company's subscriber base when existing users discontinue their subscriptions or stop using the services provided by the organization.In telecommunications, churn can affect recurring revenue, customer lifetime value, and business growth. Identifying customers who may leave can help organizations prioritize retention activities and allocate resources more effectively. Traditional churn analysis often relies on historical reports and predefined customer segments. Although useful for descriptive analysis, these methods may not fully capture complex relationships among customer demographics, service subscriptions, contract types, tenure, and billing information. Machine learning provides a data-driven approach to estimating churn probability from historical customer records. However, prediction accuracy alone is insufficient for practical use. Business users also need understandable explanations and an accessible interface for reviewing predictions. This paper introduces ChurnGuard AI, an application that combines machine-learning classification, model comparison, SHAP-based explanations, rule-based recommendations, business insights, and prediction-history storage.

The main contributions are:

1.Development of an end-to-end customer churn prediction workflow.
2. Performance Analysis of Logistic Regression, Random Forest, and XGBoost Algorithms
3.Integration of SHAP explanations for individual predictions.
4.Development of a Streamlit interface for individual and bulk customer scoring.
5.Implementation of descriptive business insights and SQLite-based prediction history.

The system supports human decision-making. It does not automatically contact customers or guarantee that recommended retention actions will prevent churn.

2. RELATED WORK

A .Machine Learning for Churn Prediction

Logistic Regression is commonly used as a baseline for binary classification because it is computationally efficient and provides a reference for evaluating more complex algorithms. This ensemble approach can model nonlinear relationships and interactions among customer attributes. XGBoost is a gradient-boosting framework that builds decision trees sequentially to reduce prediction errors. Chen and Guestrin introduced XGBoost as a scalable tree-boosting system suitable for structured datasets [1].

B. Explainable Artificial Intelligence

Complex machine-learning models can be difficult for business users to interpret. Explainable artificial intelligence techniques help describe how input features contribute to a model's output. Lundberg and Lee introduced SHAP, a framework based on Shapley values for explaining individual predictions [2]. Churn Guard AI uses SHAP to show which features contribute positively or negatively to an individual prediction.SHAP explains model behavior ; it does not prove that a particular feature causes customer churn.

C. Research Motivation

This project combines model comparison, individual prediction explanations, batch processing, and a user-facing application. Its purpose is to connect machine-learning outputs with a practical workflow for reviewing customer risk.

3. PROBLEM STATEMENT AND OBJECTIVES

A. Problem Statement

Telecommunications providers may find it difficult to identify customers who are likely to discontinue their services. A system that estimates churn probability from customer information can help staff prioritize records for further review. The problem addressed in this project is the development of an accessible application that predicts customer churn, compares machine-learning models, explains individual predictions, and presents relevant business information.

B. Objectives

The main objectives are:

1. Prepare customer data for supervised machine learning.
2. Train and compare Logistic Regression, Random Forest, and XGBoost.
3. Select a preferred model based on ROC-AUC.
4. Provide individual customer predictions and bulk CSV scoring.
5. Explain model predictions using SHAP.
6. Generate rule-based retention recommendations.
7. Display descriptive business insights.
8. Store individual prediction history using SQLite.

C. Scope

The application focuses on churn prediction and decision support using the IBM Telco Customer Churn sample dataset. Automated CRM interventions, continuous real-time event processing, experimentally verified retention ROI, and comprehensive customer-lifetime-value modelling are not verified features of the current implementation.

4. DATASET AND PREPROCESSING

A. Dataset Description

The system uses IBM's Telco Customer Churn sample dataset [3]. It contains 7,043 customer records describing demographics, tenure, subscribed services, contract arrangements, payment methods, and billing information.

Table I. Dataset Summary

Dataset characteristic	Value
Total customer records	7,043
Customers labelled as churned	1,869
Customers labelled as non-churned	5,174
Churn rate	26.54%

The churn rate represents the proportion of records labelled as churned in this dataset. It should not be generalized to every telecommunications provider.

B. Data Preprocessing

The preprocessing pipeline performs the following operations:

1. Removes customer IDbecause it is an identifier rather than a behavioural predictor.
2. Converts Total Chargesinto numeric format.
3. Converts the Yes/No churn label into a binary target.
4. Separates numeric and categorical features.
5. Fills missing numeric values using the median calculated from training data.
6. Fills missing categorical values using the most frequent training value.
7. Applies one-hot encoding to categorical features using handle_unknown='ignore'.

The fitted preprocessing pipeline is reused during prediction to maintain consistency between training and inference.

C. Training and Testing

The dataset is divided into training and testing partitions using an 80:20 ratio. Stratification preserves the target-class proportions, and random state=42makes the split reproducible. Preprocessing is fitted on the training partition and then applied to the test partition. This helps prevent information from the test data influencing the learned transformations.

5. PROPOSED METHODOLOGY

A. Logistic Regression

Logistic Regression is used as a baseline binary-classification model. It estimates the probability of a positive class using a logistic function:

$$P(y=1\mid x)=\frac{1}{1+e^{-z}}$$

Here, z represents a linear combination of input features and their learned coefficients. The reported configuration uses max iter=1500 and balanced class weights.

B. Random Forest

Random Forest combines predictions from multiple decision trees. The implemented configuration uses 200 estimators, a maximum tree depth of 12, and balanced class weights. It can represent nonlinear associations and interactions between customer attributes.

C. XGBoost

XGBoost builds decision trees sequentially, with later trees reducing errors made by the existing ensemble. The reported configuration uses:

1. 200 estimators
2. Maximum tree depth: 4
3. Learning rate: 0.05
4. Row subsampling: 0.85
5. Feature subsampling: 0.85

The preferred classifier is selected using the highest ROC-AUC among the evaluated models.

D. Evaluation Metrics

The models are evaluated using five classification metrics:

- Accuracy: Proportion of all predictions that are correct.
- Precision: Proportion of customers predicted as churners who are actually labelled as churners.
- Recall: Proportion of actual churners correctly detected.
- F1-score: Harmonic mean of precision and recall.
- ROC-AUC: Measures how effectively a model ranks positive examples above negative examples across different classification thresholds.

Because churn and non-churn records are not equally represented, accuracy should not be considered in isolation. Precision and recall provide additional information about false alarms and missed churners.

E. SHAP-Based Explanation

SHAP is integrated into the individual-customer prediction workflow. It displays feature contributions to help users understand the model output. A positive contribution generally shifts the explained output toward the positive class, while a negative contribution shifts it away. Interpretation depends on the model output and explainer configuration. The reported explanation run contained 45 transformed features. The interface displays the strongest positive and negative contributions for the selected customer. SHAP contributions are not percentages or causal effects.

F. Risk Categorization and Recommendations

The application categorizes predicted probabilities into Low, Medium, and High risk levels. The binary prediction uses a separate threshold of 0.50: probabilities at or above this threshold are labelled Churn, while lower probabilities are labelled Not Churn. Therefore, a customer may fall into a Medium risk band while receiving a Not Churn binary label because the risk band and binary prediction use different cut-offs. Retention suggestions are generated through rules based on customer attributes. These suggestions are advisory and are not automatically executed.

6. SYSTEM ARCHITECTURE AND IMPLEMENTATION

A. Architecture Overview

ChurnGuard AI consists of the following logical components:

- Data Layer: Contains the customer dataset and user-uploaded CSV files.
- Preprocessing Layer: Handles numeric and categorical imputation and one-hot encoding.
- Prediction Layer: Uses Logistic Regression, Random Forest, and XGBoost classifiers.
- Explanation Layer: Generates SHAP feature-contribution outputs.
- Interface Layer: Provides Streamlit screens for prediction, evaluation, and analysis.
- Storage Layer: Uses SQLite to store individual prediction history.

B. Application Modules



Figure 1.ChurnGuard AI Dashboard: ChurnPrediction Dashboard displaying customer totals and churn-related indicators.



Figure2.1.Single-Customer

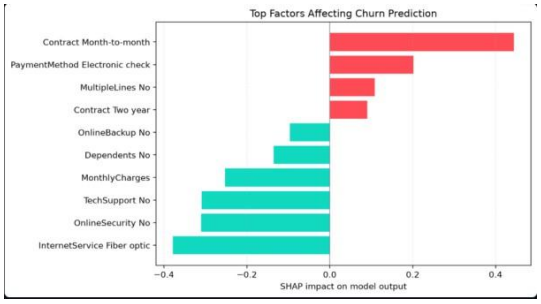


Figure2.2.Single-Customer ChurnPrediction : Individual customer inputand prediction result.

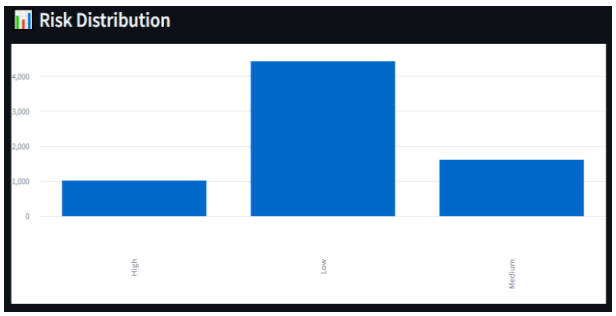


Figure 3. Bulk Customer Prediction: CSV upload and batch prediction results



.

Figure 4. Model Performance Comparison: Comparison of the three trained machine-learning classifiers.

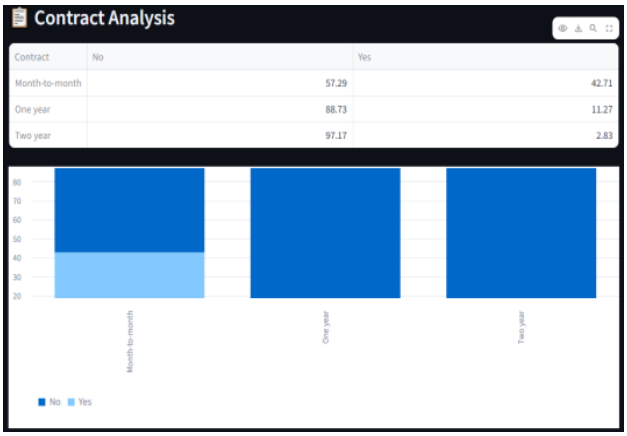


Figure 5: Business insights

C. Technology Stack

Component	Technology
Programming language	Python
Application interface	Streamlit
Data processing	Pandas, NumPy
Machine learning	Scikit-learn, XGBoost
Model explanations	SHAP
Visualization	Matplotlib and Plotly, where configured
Model persistence	Joblib
Database	SQLite
Development environment	Visual Studio Code

SQLite maintains a local record of individual prediction activity. Available stored fields include predicted probability, risk category, model label, suggested actions, predicted class, and timestamp. The current bulk-scoring workflow should not be described as saving every uploaded record to prediction history unless this behaviour has been implemented and tested.

7. RESULTS AND DISCUSSION

A. Model Performance Comparison

The following table presents the latest reported model metrics from the development training run.

Table II. Performance Comparison of Machine-Learning Models

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	73.81%	50.43%	78.88%	61.52%	84.21%
Random Forest	75.16%	52.40%	70.05%	59.95%	83.14%
XGBoost	80.06%	65.55%	52.41%	58.25%	84.43%

XGBoost achieved the highest accuracy and ROC-AUC among the three evaluated models. It is selected as the preferred model because the project's selection rule uses the highest ROC-AUC.

However, Logistic Regression achieved higher recall (78.88%) than XGBoost (52.41%). This means Logistic Regression detected a larger share of the actual churners at the reported classification threshold.

Threshold selection should be performed using validation data and a clearly defined business-cost objective, rather than repeatedly tuning against the final test set.

B. Dataset Insights

The application reports the following descriptive values:

- Total customer records: 7,043.
- Churned customers: 1,869.
- Active/non-churned customers: 5,174.
- Observed churn rate: 26.54%.
- Mean tenure: approximately 32.4 months.
- Mean monthly charge: approximately 64.76 units of the dataset currency.

The application also displays churn patterns across contract type, internet service, payment method, and tenure.

The original IBM Telco Customer Churn dataset expresses charges in US dollars. If the original values are used without conversion, they should be reported in dollars rather than rupees. Any conversion to INR must be documented.
These summaries describe associations in the sample; they do not demonstrate that a particular contract or payment method causes customers to leave.

C.SHAP Results

The following values are example contributions reported for one selected customer. They are local explanations, not global feature-importance scores.

Table III. Example SHAP Contributions

Feature	SHAP contribution
Contract - Month-to-month	+0.4443
InternetService - Fiber optic	−0.3781
OnlineSecurity - No	−0.3106
TechSupport - No	−0.3089
MonthlyCharges	−0.2527
PaymentMethod - Electronic check	+0.2015

The direction and magnitude of each contribution depend on the model output and explainer configuration. Before reporting these values, the explainer should be verified for the churn class, and the transformed feature labels should be checked against the model inputs. SHAP contributions should not be interpreted as direct causal effects.

D. Functional Testing

Manual development checks confirmed that the main application modules opened and displayed their expected outputs.

Module	Observed status
Dashboard	Displayed customer KPIs
Single Customer Prediction	Generated individual predictions
SHAP Explanation	Displayed feature-contribution information
Bulk Prediction	Accepted telco_churn.csvcontaining 7,043 records
Model Performance	Displayed model comparison and best-model ROC-AUC
Business Insights	Displayed dataset statistics and segment analysis
Prediction History	Displayed saved individual prediction records
About	Displayed project description and technology stack

These observations indicate that the basic workflows operated during development. They do not replace automated tests, independent validation, or production-scale testing.

9. LIMITATIONS

The current implementation has several limitations:

- Single dataset: Evaluation uses one public sample dataset. Performance may differ for other providers, customer populations, and time periods.
- Single train-test split: The reported results are based on one stratified split. Repeated cross-validation and temporal evaluation have not been confirmed.

- Recall trade-off: The selected XGBoost model reports 52.41% recall, meaning it misses a substantial share of actual churners at the reported threshold.
- Explanation limitations: SHAP describes model outputs but does not establish cause-and-effect relationships.
- Descriptive analytics: The current analysis focuses on customer counts and service-related segments. Comprehensive cohort analysis and customer-lifetime-value modelling are not verified as implemented.
- No automatic intervention: Users review the suggested retention actions; the application does not automatically contact customers.
- No measured retention ROI: No controlled experiment has established that the recommendations reduce churn or generate a specific financial return.
- Deployment and security: Authentication, database safeguards, and deployment configuration require further review before public use.

10. CONCLUSION AND FUTURE WORK

This paper presented ChurnGuard AI, an interactive customer churn prediction and retention-support application developed using Python, Streamlit, Scikit-learn, XGBoost, SHAP, and SQLite.

The application combines data preprocessing, comparative model evaluation, individual and bulk predictions, explainable AI, descriptive business insights, rule-based recommendations, and individual prediction-history storage.

In the latest reported training run, XGBoost achieved 80.06% accuracy and 84.43% ROC-AUC, while Logistic Regression achieved higher recall. These results show that the model with the highest ROC-AUC does not necessarily detect the largest proportion of churners at the selected classification threshold.

The system can support exploratory customer-risk assessment, but its predictions should be treated as decision-support information rather than definitive statements about future customer behaviour.

Future work includes repeated cross-validation, temporal and external validation, probability calibration, validation-based threshold optimization, cost-sensitive evaluation, stronger authentication, automated testing, and cloud deployment. More advanced cohort analysis, customer-lifetime-value estimation, and CRM integration can also be developed and evaluated as future extensions.

REFERENCES

1) Guan, Z. (2025). Scalable framework for basketball game prediction combining image processing XGBoost and enhanced support vector machine. Scalable Computing: Practice and Experience, 26(4). https://doi.org/10.12694/scpe.v26i4.4684

2) Sturmfels, P., Lundberg, S., & Lee, S.-I. (2020). Visualizing the impact of feature attribution baselines. Distill, 5(1). https://doi.org/10.23915/distill.00022

3) IBM. (2024). GitHub - IBM/telco-customer-churn-on-ICP4D: Data analysis, model building and deploying with WML on IBM cloud pak for data. GitHub. https://github.com/IBM/telco-customer-churn-on-icp4d

4) van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., & Yu, T. (2014). Scikit-image: Image processing in Python. PeerJ, 2, e453. https://doi.org/10.7717/peerj.453

5) Welcome to the SHAP documentation — SHAP latest documentation. (n.d.).Shap.Readthedocs.Io. Retrieved November 3, 2026, from https://shap.readthedocs.io/

6) Streamlit. (2024). Streamlit docs. Docs.Streamlit.Io. https://docs.streamlit.io/

7) SQLite. (2019). SQLite documentation. Sqlite.Org. https://www.sqlite.org/docs.html

BIOGRAPHIES

1) PROF. J.S.Jadhav

FACULTY OF DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING INTERNATIONAL CENTER OF EXCELLENCE IN ENGGINERING & MANAGEMENT (ENGINEERING)

2) Shaikh Zaid

PURSUING BACHELOR OF ENGINEERING (B.E.) IN COMPUTER SCIENCE AND ENGINEERING (CSE) AT INTERNATIONAL CENTER OF EXCELLENCE IN ENGINEERING AND MANAGEMENT (ICEEM).

3) Yash puse

PURSUING BACHELOR OF ENGINEERING (B.E.) IN COMPUTER SCIENCE AND ENGINEERING (CSE) AT INTERNATIONAL CENTER OF EXCELLENCE IN ENGINEERING AND MANAGEMENT (ICEEM).

4) Gaurav shedge

PURSUING BACHELOR OF ENGINEERING (B.E.) IN COMPUTER SCIENCE AND ENGINEERING (CSE) AT INTERNATIONAL CENTER OF EXCELLENCE IN ENGINEERING AND MANAGEMENT (ICEEM).