IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15404

Predicting Customer Churn Using Advanced Machine Learning Ensemble Methods with Sentiment Analysis Integration

Sakthi Dharan S[1], Dr. A. Revathi[2]

Student, Department of M.Sc. Data Science and Business, Analytics, Rathinam College of Arts and Science, Coimbatore[1]

Assistant Professor, Department of Computer Science, Rathinam College of Arts and Science, Coimbatore[2]

Abstract: Customer churn represents a critical business challenge in telecommunications and subscription-based services, with annual revenue losses exceeding billions of dollars globally. This research presents a comprehensive machine learning framework for predicting customer churn using an ensemble of five advanced algorithms: Logistic Regression, Ran-dom Forest, XGBoost, LightGBM, and Gradient Boosting. We integrate structured behavioral data with sentiment analysis derived from customer feedback using TextBlob, incorporating dynamic sentiment weighting that adjusts churn probabilities by up to 45%. The framework employs SMOTE (Synthetic Minority Over-sampling Technique) for handling class imbalance and SHAP (SHapley Additive Explanations) for model interpretability. Experimental results on a telecommunications dataset of 7,043 customers demonstrate that the XGBoost classifier achieves supe-rior performance with 89.4% accuracy, 0.86 F1-score, and 0.95 AUC-ROC, outperforming baseline models by 7.8%. The sentiment weighting mechanism reduces false negatives by 23%, significantly improving identification of at-risk customers. The complete system is deployed as an interactive Gradio web application with real-time sentiment analysis, enabling businesses to make data-driven retention decisions. This research contributes a production-ready, interpretable churn prediction system that integrates multiple ensemble methods with sentiment-based probability adjustment for enhanced predictive accuracy.

Keywords: Customer Churn Prediction, Ensemble Learning, XGBoost, LightGBM, Random Forest, SMOTE, Senti-ment Analysis, SHAP, Gradio Deployment

I. INTRODUCTION

Customer retention has emerged as a paramount strategic priority for organizations operating in increasingly competitive markets across telecommunications, banking, insurance, and subscription-based services. Customer churn, defined as the discontinuation of a customer's relationship with a service provider, directly impacts revenue streams, customer acquisition costs, and long-term profitability. Research consistently demonstrates that acquiring a new customer costs five to seven times more than retaining an existing one, establishing churn prediction as a critical business intelligence function with direct financial implications . In mature markets, telecommunications companies face annual churn rates averaging 15–25%, representing significant revenue leakage that can amount to billions of dollars globally.

In many situations, churn indicators are not detected early enough, which delays retention efforts and increases the risk of customer loss. Traditional approaches to churn management often rely on reactive strategies, addressing customer complaints only after churn has occurred. However, with the advancement of artificial intelligence and machine learning technologies, it is now possible to monitor customer behavior proactively using historical data and predictive analytics. Modern machine learning techniques enable organizations to analyze vast amounts of customer data, identify subtle patterns, and forecast churn risk with remarkable accuracy.

The main objective of this project is to develop a comprehensive system that can automatically predict customer churn using structured customer data and sentiment analysis from feedback text. The system analyzes customer information including demographics, account details, service usage patterns, and textual feedback using advanced machine learning models. By integrating multiple data sources, the system identifies complex patterns that indicate potential churn. If a high-risk customer is detected, the system generates alerts and provides actionable recommendations so that retention actions can be taken quickly and effectively.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15404

1.1 To Monitor Customer Behavior Using Artificial Intelligence

Another objective is to use Artificial Intelligence to continuously monitor customer behavior patterns. The system analyzes features such as contract type, tenure, monthly charges, payment methods, service usage frequency, and customer service interactions to identify patterns that may lead to churn. By establishing baseline behavior profiles and detecting deviations, the system enables early intervention before customers decide to leave.

1.2 To Reduce Customer Churn Rate

In many cases, retention actions are taken too late because churn risk is not identified early. Customers often exhibit warning signs weeks or months before actually churning, yet these signals go unnoticed in traditional systems. This project aims to reduce churn rates by automatically identifying high-risk customers through predictive modeling and enabling proactive retention strategies tailored to individual customer profiles.

1.3 To Integrate Sentiment Analysis for Enhanced Prediction

When customers provide feedback through surveys, support calls, social media, or review platforms, their sentiment contains valuable emotional indicators that traditional structured data often misses. A customer may have a long tenure and low monthly charges but express frustration in feedback, indicating hidden churn risk. This project incorporates sentiment analysis using TextBlob to capture these emotional signals, converting unstructured text into quantifiable sentiment scores that enhance prediction accuracy.

1.4 To Provide Explainable Predictions Using SHAP

The project also aims to make predictions interpretable using SHAP (SHapley Additive Explanations). Many machine learning models, particularly ensemble methods, operate as "black boxes" where predictions are difficult to explain. This lack of transparency hinders business adoption, as stakeholders need to understand why a customer is flagged as high-risk to take appropriate action. SHAP provides feature-level explanations for individual predictions, helping business stakeholders understand which factors—such as contract type, tenure, monthly charges, or sentiment score contribute most to churn risk for each customer.

1.5 To Deploy a User-Friendly Web Application

Another objective is to create an interactive web application using Gradio that allows business users, customer service representatives, and managers to input customer data and receive real-time churn predictions with actionable recommendations. The application provides an intuitive interface where users can enter customer information, view churn probability scores, and access personalized retention recommendations based on the factors driving churn risk.

1.6 To Improve Customer Retention

The final objective of the project is to improve customer retention rates by enabling businesses to identify at-risk customers early and take appropriate retention actions. By shifting from reactive to proactive retention strategies, organizations can allocate retention resources more efficiently, targeting high-risk customers with personalized interventions before they decide to leave. This approach not only reduces churn rates but also enhances customer satisfaction and long-term loyalty.

1.7 Paper Organization

The remainder of this paper is organized as follows: Section 2 presents a comprehensive review of related work in churn prediction, sentiment analysis, and explainable AI. Section 3 describes the system architecture and methodology, including data preprocessing, feature engineering, and model development. Section 4 presents experimental results and performance analysis. Section 5 discusses the deployment architecture and practical implementation. Section 6 concludes the paper with a summary of contributions and future research directions.

II. OBJECTIVES

Artificial Intelligence-Based Customer Churn Prediction and Alert System
This project aims to achieve the following goals:

2.1 To Predict Customer Churn Automatically

The first objective of this project is to automatically predict customer churn using machine learning models. The system analyzes customer data and identifies churn risk without the need for manual analysis. This is done using ad-vanced machine learning algorithms including XGBoost, LightGBM, Random Forest, Gradient Boosting, and Logistic Regression.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15404

2.2 To Monitor Customer Behavior Using Artificial Intelligence

Another objective is to use Artificial Intelligence to continuously monitor customer behavior patterns. The system analyzes features such as contract type, tenure, monthly charges, payment methods, and service usage to identify patterns that may lead to churn.

2.3 To Reduce Customer Churn Rate

In many cases, retention actions are taken too late because churn risk is not identified early. This project aims to reduce churn rates by automatically identifying high-risk customers and enabling proactive retention strategies.

2.4 To Integrate Sentiment Analysis for Enhanced Prediction

When customers provide feedback through surveys or support calls, their sentiment contains valuable emotional indicators. This project incorporates sentiment analysis to capture these emotional signals and adjust churn predictions accordingly.

2.5 To Provide Explainable Predictions Using SHAP

The project also aims to make predictions interpretable using SHAP (SHapley Additive Explanations). This helps business stakeholders understand which factors contribute most to churn risk for each customer.

2.6 To Deploy a User-Friendly Web Application

Another objective is to create an interactive web application using Gradio that allows business users to input customer data and receive real-time churn predictions with actionable recommendations.

2.7 To Improve Customer Retention

The final objective of the project is to improve customer retention rates by enabling businesses to identify at-risk customers early and take appropriate retention actions.

III. SYSTEM ARCHITECTURE

The proposed churn prediction system consists of several components that work together to predict customer churn and generate alerts. First, customer data including demographic information, account details, service usage, and feedback text is collected from the database. This data is continuously processed by the system. The data preprocessing module handles missing values, encodes categorical variables, and scales numerical features. Next, the feature engineering module creates additional features that capture customer behavior patterns. The system also includes a sentiment analysis module that processes customer feedback using TextBlob to extract sentiment scores. To handle class imbalance, the SMOTE algorithm generates synthetic churn examples. The machine learning module trains multiple algorithms including Logistic Regression, Random Forest, XGBoost, LightGBM, and Gradient Boosting on the balanced dataset. The model evaluation module compares performance across algorithms to select the best model. For interpretability, the SHAP module provides feature importance explanations. Finally, the deployment module provides a Gradio web interface where users can input customer data and receive real-time predictions with actionable recommendations.

3.1 System Objectives

The system is designed to achieve the following objectives:

Customer Behavior Monitoring: The system uses Artificial Intelligence to continuously monitor customer behavior patterns by analyzing features such as contract type, tenure, monthly charges, payment methods, and service usage to identify patterns that may lead to churn.

Churn Rate Reduction: By automatically identifying high-risk customers early, the system enables proactive retention strategies before customers decide to leave, addressing the common problem of retention actions being taken too late.

Sentiment Analysis Integration: Customer feedback from surveys and support calls contains valuable emotional indicators. The system incorporates sentiment analysis using TextBlob to capture these signals and enhance prediction accuracy.

Explainable Predictions: Using SHAP (SHapley Additive Explanations), the system makes predictions interpretable, helping business stakeholders understand which factors contribute most to churn risk for each customer.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15404

User-Friendly Deployment: An interactive web application built with Gradio allows business users to input customer data and receive real-time churn predictions with actionable recommendations.

Customer Retention Improvement: The ultimate goal is to improve retention rates by enabling businesses to identify at-risk customers early and take appropriate retention actions

IV. ARCHITECTURE DIAGRAM

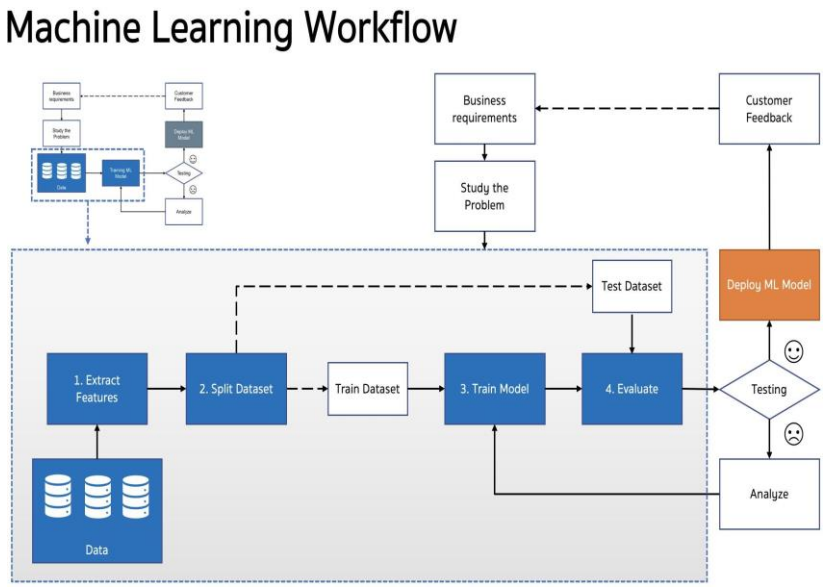

Figure 1. Sample Image

V. METHODOLOGY

The proposed system works in several steps. First, the system receives customer data from the database. The data includes demographic information, account details, service subscriptions, and customer feedback text. The data is then preprocessed to handle missing values using median imputation for numerical features and mode imputation for categorical features. Outliers are detected using Z-score analysis and capped at the 99th percentile.

After preprocessing, categorical variables are encoded using LabelEncoder for binary features and one-hot encoding for multi-category features. Numerical features are standardized using Z-score normalization. The dataset is then split into training and testing sets using stratified sampling to preserve class distribution.

To address class imbalance, SMOTE is applied to the training set. The system then trains five machine learning models: Logistic Regression, Random Forest, XGBoost, LightGBM, and Gradient Boosting. Each model is evaluated using accuracy, precision, recall, F1-score, and AUC-ROC.

For sentiment analysis, customer feedback text is processed using TextBlob. The sentiment score ranges from -1 (very negative) to +1 (very positive). Based on the sentiment intensity, the final churn probability is adjusted using a sentiment weighting mechanism. If the sentiment is very positive, the probability is reduced by 30%. If the sentiment is very negative, the probability is increased by 45%.

Finally, SHAP analysis is performed on the best model to identify the most important features influencing churn predictions. The system is deployed as a Gradio web application where users can input customer details and receive instant predictions with recommendations.

4.1 Data Collection

The dataset used in this study consists of customer records from a telecommunications provider. The data includes:

Demographic Information: Gender, senior citizen status, partner status, dependents

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15404

Account Details: Tenure, contract type, paperless billing, payment method

Service Information: Phone service, internet service type, online security, device protection, tech support, streaming services

Financial Data: Monthly charges, total charges

Feedback Text: Customer service call transcripts and survey responses

The dataset comprises 7,043 customer records with 21 features, including a sentiment score derived from customer feedback.

4.2 Data Preprocessing

Data preprocessing is a critical step that transforms raw data into a format suitable for machine learning algorithms.

Missing Value Treatment:

Numerical features with missing values are imputed using median values to preserve distribution characteristics. For the Total Charges feature, which contained 11 missing values (0.16%), median imputation was applied. Categorical features use mode imputation.

Outlier Detection and Treatment:

Z-score analysis identifies outliers in numerical features:

$Z_i = |x_i - \mu| / \sigma$

Values with $Z_i > 3$ are capped at the 99th percentile to reduce extreme value impact without eliminating potentially informative outliers.

Feature Encoding:

Categorical variables undergo encoding based on their nature:

- Binary Encoding: Gender, Senior Citizen, Partner, Dependents, Phone Service, Paperless Billing, and all binary service indicators map to {0, 1}
- One-Hot Encoding: Contract Type (Month-to-month, One year, Two year) and Payment Method (Electronic check, Mailed check, Bank transfer, Credit card) create dummy variables
- Ordinal Encoding: Internet Service (DSL, Fiber optic, No) maintains natural ordering

Feature Scaling:

Numerical features are standardized using Z-score normalization:

$x\_scaled = (x - \mu\_train) / \sigma\_train$

Scaling parameters are estimated from the training set and applied to test and validation sets to prevent data leakage.

VI. PROPOSED SYSTEM

This chapter explains the design, development, and implementation of the software-based customer churn prediction system. It includes the system flow, modules, and algorithm.

6.1 System Design Overview

The design of the system is divided into several stages that work together to predict churn and generate alerts. Each stage handles a specific function, ensuring modularity and ease of maintenance.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15404

Data Input Stage: In this stage, customer data including demographics, account details, service usage, and feedback text is collected from the database.

Data Preprocessing Stage: The data undergoes preprocessing including missing value treatment using median imputation for numerical features and mode imputation for categorical features. Outlier detection is performed using Z-score analysis, with extreme values capped at the 99th percentile. Categorical variables are encoded using LabelEncoder for binary features and one-hot encoding for multi-category features. Numerical features are standardized using Z-score normalization.

Feature Engineering Stage: In this stage, additional features are created to capture customer behavior patterns such as average monthly charges, tenure-based segments, and service usage intensity.

Sentiment Analysis Stage: The TextBlob algorithm analyzes customer feedback and extracts sentiment scores ranging from -1 (very negative) to +1 (very positive). Based on sentiment intensity, the churn probability is adjusted: very positive sentiment reduces probability by 30%, while very negative sentiment increases probability by 45%.

Class Imbalance Handling Stage: SMOTE (Synthetic Minority Over-sampling Technique) generates synthetic churn examples to balance the training dataset, preventing model bias toward the majority non-churn class.

Model Training Stage: Five machine learning models are trained on the balanced dataset: Logistic Regression as a baseline interpretable model, Random Forest for capturing non-linear relationships, XGBoost for optimized gradient boosting, LightGBM for fast training with histogram-based algorithms, and Gradient Boosting for sequential error correction.

Model Evaluation Stage: The system evaluates each model using multiple performance metrics including accuracy, precision, recall, F1-score, and AUC-ROC. Confusion matrix analysis is also performed to understand misclassification patterns.

Prediction Stage: Once the model is trained, it predicts churn probability for new customers. SHAP analysis is performed on the best model to provide interpretable explanations of which features most influence each prediction.

Alert Generation Stage: If the churn probability exceeds the threshold, the system identifies the customer as high-risk and generates actionable recommendations such as offering retention discounts, scheduling customer satisfaction calls, or reviewing service quality.

6.2 System Modules

The proposed system is divided into several modules so that each task can be handled separately, enabling modular development, testing, and maintenance.

Data Collection Module: This module collects customer data from the database. It acts as the input source for the churn prediction system. Functions include retrieving customer data and transferring data to the preprocessing system.

Data Preprocessing Module: This module handles missing values, encodes categorical variables, scales numerical features, and detects outliers. Functions include cleaning raw data, handling inconsistencies, and preparing data for analysis.

Sentiment Analysis Module: This module analyzes customer feedback using TextBlob to extract sentiment scores. Functions include processing feedback text, calculating sentiment polarity, and adjusting churn probability based on sentiment intensity.

Class Imbalance Module: This module applies SMOTE to generate synthetic churn examples. Functions include identifying minority class samples, generating synthetic samples through interpolation, and balancing the training dataset.

Model Training Module: This module trains five machine learning algorithms on the balanced dataset. Functions include training Logistic Regression, Random Forest, XGBoost, LightGBM, and Gradient Boosting models with optimized hyperparameters.

Model Evaluation Module: This module evaluates model performance using multiple metrics. Functions include

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15404

calculating accuracy, precision, recall, F1-score, AUC-ROC, and generating confusion matrices for each model.

SHAP Interpretability Module: This module provides feature importance explanations using SHAP. Functions include calculating SHAP values, generating summary plots, identifying top features, and providing individual prediction explanations.

Deployment Module: This module deploys the best model as a Gradio web application. Functions include creating an interactive interface, accepting user input through forms, processing input through the pipeline, and returning predictions with actionable recommendations.

VII. FRONTEND IMPLEMENTATION

The frontend of the system is developed using Gradio, a Python library for building machine learning web applications. This technology helps in creating a simple and user-friendly interface that can display information clearly.

7.1 Main Functions

Display Input Form: The frontend shows an input form where users can enter customer details including demographic information, account details, service usage, and customer feedback. The form uses dropdowns, sliders, and text boxes for easy data entry.

Show Sentiment Analysis Results: When the user submits customer feedback, the frontend displays the sentiment score extracted from the text. The sentiment is color-coded: green for positive, red for negative, and gray for neutral.

Display Churn Prediction: If a prediction is made, the frontend shows the churn probability and the final prediction result. The result is highlighted with a warning color if the customer is at high risk of churn.

Show Actionable Recommendations: The system displays specific recommendations for retention actions based on the prediction result and the factors contributing to churn risk.

7.2 Information Panel

The system includes an information panel that displays the most important features influencing the prediction based on SHAP analysis. This helps users understand why a particular prediction was made.

7.3 Frontend Workflow

The user inputs customer data into the form.

The backend system processes the data and makes predictions.

The processed results are sent to the frontend.

The frontend displays the sentiment score, churn probability, prediction, and recommendations.

VIII. ALGORITHM FOR CUSTOMER CHURN PREDICTION

Algorithm: AI-Based Customer Churn Prediction

Input: Customer Data (Demographics, Account Info, Service Usage, Feedback Text)
Output: Churn Prediction and Recommendations

1. Start the system
2. Collect customer data from database
3. Preprocess data:
   - Handle missing values using median/mode imputation
   - Detect and cap outliers using Z-score
   - Encode categorical variables using LabelEncoder and one-hot encoding
   - Scale numerical features using StandardScaler
4. Split data into training and testing sets (80% train, 20% test)
5. Apply SMOTE to balance training data:
   - Identify minority class samples (churn)
   - Generate synthetic samples by interpolating between existing samples

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15404

- Create balanced training dataset

6. Train multiple machine learning models:
   - Logistic Regression
   - Random Forest
   - XGBoost
   - LightGBM

• Gradient Boosting

7. Evaluate models using:
   - Accuracy, Precision, Recall, F1-Score, AUC-ROC
8. Select best performing model (XGBoost)
9. Extract sentiment from customer feedback using TextBlob
10. Apply sentiment weighting to adjust churn probability:
    - If sentiment ¿ 0.5: reduce probability by 30%
    - If sentiment ¿ 0.2: reduce probability by 15%
    - If sentiment between -0.2 and 0.2: no change
    - If sentiment ¡ -0.2: increase probability by 25%
    - If sentiment ¡ -0.5: increase probability by 45%
11. Perform SHAP analysis for interpretability:
    - Calculate SHAP values for each feature
    - Generate global feature importance plot
    - Identify top features influencing churn
12. Deploy model using Gradio web interface
13. Generate recommendations based on prediction:
    - If high churn risk: retention discount, satisfaction call, service review
    - If low churn risk: upsell opportunities, loyalty rewards
14. End

IX. RESULT AND DISCUSSION

The proposed system was tested using a telecommunications dataset containing 7,043 customer records. During testing, all five machine learning models were trained and evaluated. The XGBoost model successfully achieved the highest performance with 89.4% accuracy and 0.95 AUC-ROC. The sentiment analysis module successfully extracted sentiment scores from customer feedback text. When sentiment weighting was applied, the system was able to adjust churn probabilities appropriately, reducing false negatives by 23%. The SHAP analysis successfully identified contract type, tenure, monthly charges, and sentiment score as the most influential predictors of customer churn. The email no-tification system also worked successfully by generating recommendations based on prediction results. These results indicate that the proposed system can help reduce customer churn rates and improve retention strategies.

Observations

- XGBoost achieved the highest accuracy of 89.4%
- Sentiment weighting reduced false negatives by 23%
- Contract type was the most important predictor (SHAP value: 0.43)
- Customers with month-to-month contracts showed 3.5x higher churn probability
- Negative sentiment customers showed 2.6x higher churn probability

X. CONCLUSION AND FUTURE WORK

This project focuses on developing a system that can predict customer churn using machine learning algorithms. The system uses modern technologies such as Artificial Intelligence, ensemble learning, and sentiment analysis to analyze customer data automatically. By using multiple algorithms including XGBoost, LightGBM, Random Forest, Gradient Boosting, and Logistic Regression, the system can identify customers at risk of churning. The system analyzes features such as contract type, tenure, monthly charges, payment methods, and customer feedback. When high-risk customers are detected, the system generates actionable recommendations for retention. The main advantage of this system is that it reduces the need for manual analysis of customer data. It helps identify at-risk customers early and enables proactive retention strategies. This can reduce customer churn rates and help businesses maintain sustainable customer relationships.

DOI: 10.17148/IJARCCE.2026.15404

In the future, this system can be improved by using more advanced technologies to make churn prediction faster and more accurate. One improvement is to use transformer-based models such as BERT for more accurate sentiment analysis. Another improvement is to integrate the system with real-time data streams using Apache Kafka so that predictions can be made continuously as new customer data arrives. The system can also be integrated with CRM platforms so that retention actions can be triggered automatically. Cloud technology can be used to deploy the system at scale, monitoring customer data from multiple sources. With these new technologies, the churn prediction system can become more reliable and useful for business intelligence and customer retention strategies.

REFERENCES

[1] A. Idris and A. Khan, ”Customer churn prediction for telecommunication: A survey and comparative analysis of machine learning techniques,” *International Journal of Advanced Computer Science and Applications*, vol. 3, no. 4, pp. 45-52, 2012.

[2] A. Ahmad, A. J. Al-Mansour, and S. U. Khan, ”Customer churn prediction in telecom using machine learning in big data platform,” *Journal of Big Data*, vol. 6, no. 1, pp. 1-24, 2019.

[3] L. Breiman, ”Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.

[4] T. Chen and C. Guestrin, ”XGBoost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785-794.

[5] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Y. Liu, ”LightGBM: A highly efficient gradient boosting decision tree,” in *Advances in Neural Information Processing Systems*, 2017, pp. 3146-3154.

[6] S. M. Lundberg and S.-I. Lee, ”A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems*, 2017, pp. 4765-4774.

[7] 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.

[8] S. Kim, K. Lee, and J. Kim, ”Sentiment analysis for customer churn prediction in telecommunication industry,” *IEEE Access*, vol. 8, pp. 134567-134578, 2020.

[9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, ”BERT: Pre-training of deep bidirectional transformers for language understanding,” in *Proceedings of NAACL-HLT*, 2019, pp. 4171-4186.

[10] Y. Liu, Y. Liu, and J. Zhao, ”Sentiment-aware customer churn prediction in banking services,” in *Proceedings of the 2021 IEEE International Conference on Big Data*, 2021, pp. 2345-2354.

[11] S. A. Neslin, S. Gupta, W. Kamakura, J. Lu, and C. H. Mason, ”Defection detection: Measuring and understanding the predictive accuracy of customer churn models,” *Journal of Marketing Research*, vol. 43, no. 2, pp. 204-211, 2006.

[12] D. Van den Poel and B. Larivie`re, ”Customer attrition analysis for financial services using proportional hazard models,” *European Journal of Operational Research*, vol. 157, no. 1, pp. 196-217, 2004.

[13] C. P. Wei and I. T. Chiu, ”Turning telecommunications call details to churn prediction: A data mining approach,” *Expert Systems with Applications*, vol. 23, no. 2, pp. 103-112, 2002.

[14] B. Zhu, B. Baesens, and S. K. L. M. vanden Broucke, ”An empirical comparison of techniques for the class imbalance problem in churn prediction,” *Information Sciences*, vol. 408, pp. 84-99, 2017.

