ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2026, Volume 7 Issue 2
www.ijisea.org

## Article Info

Date Received: 15/03/2026
Date Revised: 05/04/2026
Available Online: 27/04/2026

# Explainable Retention Intelligence Framework for Telecom Customer Attrition Prediction Using Ensemble Learning and Interactive Business Analytics

**1. J. Rakesh, 2. K. Akash, 3. K. Vasanthi, 4. K. Naga Anjali Devi, 5. K. Sai Kiran, 6. S. Mohan Babu Chowdary**

**Author Affiliations**

**1,2,3,4,5. B. Tech CSE (AIML) Students, Dept. of CSE, Sir C R Reddy College of Engineering, Eluru.**
**6. Assistant Professor, Dept. of CSE, Sir C R Reddy College of Engineering, Eluru.**

DOI: 10.64264/ijisea/0740

## ABSTARCT

Customer attrition prediction is a critical challenge in the telecommunications industry due to its direct impact on customer retention and organizational revenue. Traditional churn analysis approaches primarily rely on descriptive statistics and reactive strategies, which are insufficient for identifying high-risk customers at an early stage. This paper proposes an explainable retention intelligence framework for telecom customer attrition prediction using ensemble learning techniques and interactive business intelligence visualization.

The proposed framework applies supervised machine learning algorithms including Logistic Regression, Random Forest, and XGBoost on a publicly available telecom customer churn dataset obtained from Kaggle containing demographic characteristics, service usage behavior, and billing attributes. Among these models, the Random Forest classifier achieved superior predictive performance with an accuracy of 0.7660 and ROC–AUC score of 0.8440, demonstrating effective capability in identifying customers at risk of attrition.

To enhance interpretability and decision-making support, explainable artificial intelligence techniques such as SHAP and LIME were employed to identify the key factors influencing customer attrition. The prediction results were further integrated with interactive Power BI dashboards to visualize churn-risk segments and retention insights, enabling proactive and data-driven retention strategy development for telecom service providers.

Key words: Customer Attrition Prediction, Telecom Customer Churn, Ensemble Learning, Explainable Artificial Intelligence (XAI), Random Forest Classifier.

## 1. INTRODUCTION

The telecommunications industry is essential for global connectivity, offering voice, internet, and digital services. With rapid technological growth, competition among telecom providers has increased, making customer retention crucial for maintaining profitability and market share.Customer churn, where users switch providers, poses a major challenge as it impacts revenue and customer lifetime value.

Since acquiring new customers is costlier than retaining existing ones, predicting churn in advance has become a strategic priority. Traditional systems rely on descriptive and reactive approaches, which are not sufficient for accurate prediction.

Machine learning techniques provide effective solutions by analyzing customer data such as demographics, usage patterns, and billing information. Models like Logistic Regression, Random Forest, and XGBoost are commonly used due to their ability to handle structured data and complex relationships.

However, many existing models focus only on accuracy and lack interpretability and business insights. To address this, integrating Explainable AI techniques like SHAP and LIME helps identify key factors influencing churn. Additionally, combining predictive models with Power BI dashboards enables interactive visualization and supports data-driven decision-making. This study proposes an integrated framework using ensemble learning, explainable AI, and visualization tools to identify high-risk customers and improve retention strategies.

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2026, Volume 7 Issue 2
www.ijisea.org

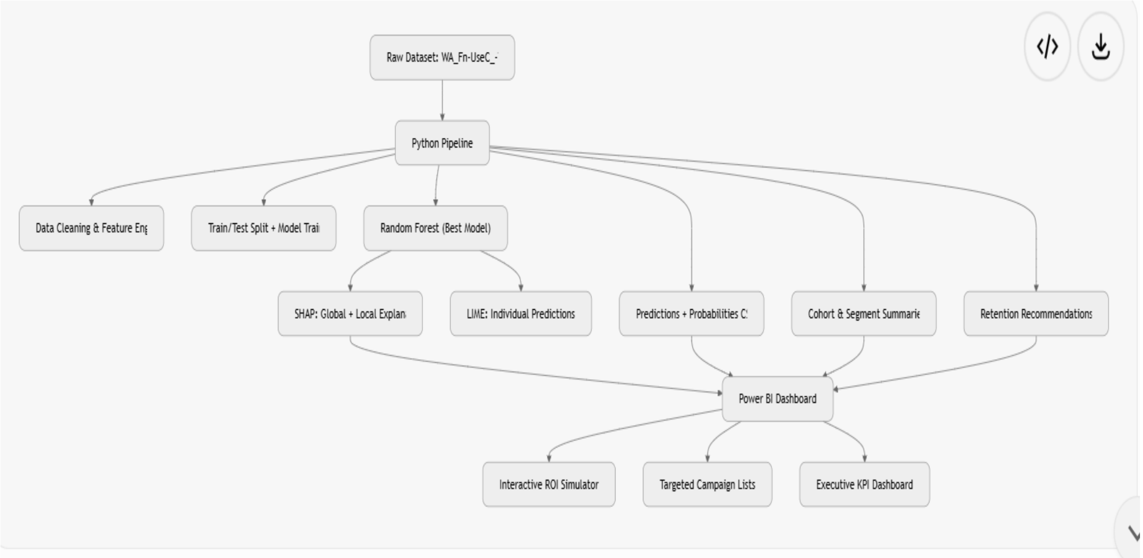



*Fig. 1 Proposed Customer Churn Prediction Framework Architecture*

The main contributions of this study are summarized as follows:

- Development of an ensemble learning-based churn prediction framework using Logistic Regression, Random Forest, and XGBoost.
- Application of SHAP and LIME techniques to improve interpretability and identify key churn-driving factors.
- Integration of predictive analytics with interactive Power BI dashboards for visualization of churn-risk insights.
- Transformation of traditional churn analysis into a proactive retention intelligence framework supporting data-driven telecom decision-making.

## 2. LITERATURE REVIEW

Customer churn prediction is a key research area in customer relationship management and predictive analytics, especially in telecom industries where competition is high. Organizations increasingly use data-driven methods to identify customers likely to leave and to design effective retention strategies.

Early approaches used traditional statistical methods like Logistic Regression and survival analysis, which provided basic insights but struggled with complex relationships in data. With the growth of large datasets, machine learning techniques such as decision trees, support vector machines, and neural networks improved prediction accuracy.Ensemble methods like Random Forest and XGBoost further enhanced performance by capturing complex feature interactions.

However, accuracy alone is not sufficient for business decision-making.Recent research highlights the importance of interpretability. Techniques like SHAP and LIME help explain model predictions by identifying key influencing factors. Additionally, integrating predictive results with tools like Microsoft Power BI enables better visualization of churn trends and customer insights. This study addresses the gap by combining ensemble models, explainable AI, and BI dashboards into a unified churn prediction framework.

| Study | Method | Strength | Limitation |
|---|---|---|---|
| [8] Verbeke et al. | Rule-based churn models | Interpretability | Lower predictive accuracy |
| [9] Hadden et al. | Data mining churn systems | Framework-level analysis | Limited deployment integration |
| [3] Breiman | Random Forest | High accuracy | Limited explainability |
| [6] Lundberg & Lee | SHAP | Model transparency | No business visualization integration |
| [7] Ribeiro et al. | LIME | Local explanation capability | No telecom deployment pipeline |
| Proposed Work | RF + SHAP + LIME + Power BI | Prediction + Explainability + BI | - |

### 2.1 Critical Analysis and Research Gap

Many machine learning models have been used for telecom churn prediction, but most focus mainly on improving accuracy while neglecting interpretability and business-level decision support. Traditional models like Logistic Regression are easy to interpret but cannot effectively capture complex relationships, whereas advanced models like Random Forest and gradient boosting offer better performance but act as black-box systems with limited transparency.
Although Explainable AI techniques such as SHAP and LIME improve model transparency, their integration with business intelligence platforms is still limited. There is a lack of unified frameworks that combine prediction, explainability, and interactive visualization for effective decision-making. This study addresses the gap by integrating ensemble models, SHAP–LIME techniques, and Microsoft Power BI into a single churn analytics framework.

## 3. METHODOLOGY

This section describes the systematic methodology adopted to develop the customer churn prediction framework for the telecommunications industry. The proposed approach integrates data preprocessing, supervised machine learning classification models, explainable artificial

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2026, Volume 7 Issue 2
www.ijisea.org

intelligence techniques, and business intelligence visualization to identify customers at risk of discontinuing telecom services. The overall workflow follows a structured data mining lifecycle based on the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology to ensure reliability, reproducibility, and interpretability of predictive analytics results.

The research workflow followed in this study is illustrated in Fig. 2, which represents the CRISP-DM lifecycle used to guide the development of the churn prediction framework.

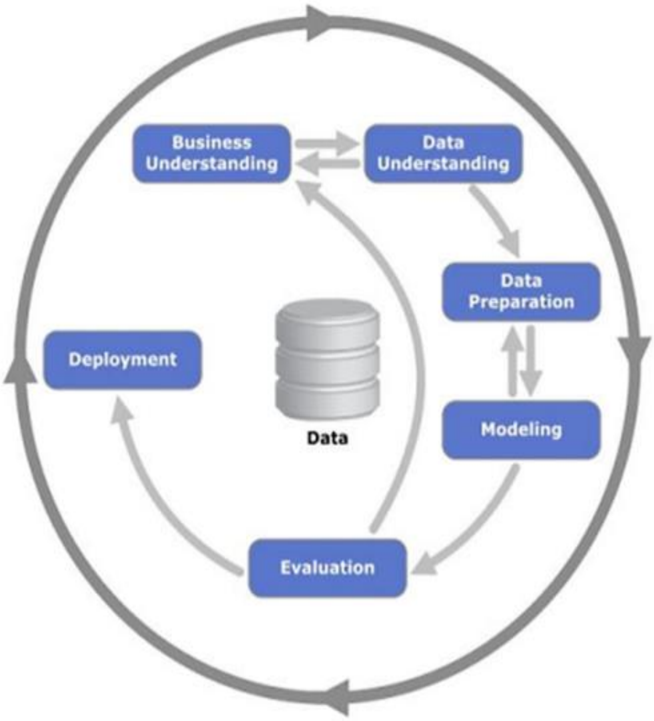



*Fig. 2 CRISP-DM methodology followed in the proposed churn prediction framework*

Furthermore, the detailed implementation pipeline of the proposed churn prediction system is illustrated in **Fig. 3**, showing the sequence of preprocessing, feature engineering, model training, evaluation, explainability analysis, and visualization stages used in the framework.

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2026, Volume 7 Issue 2
www.ijisea.org

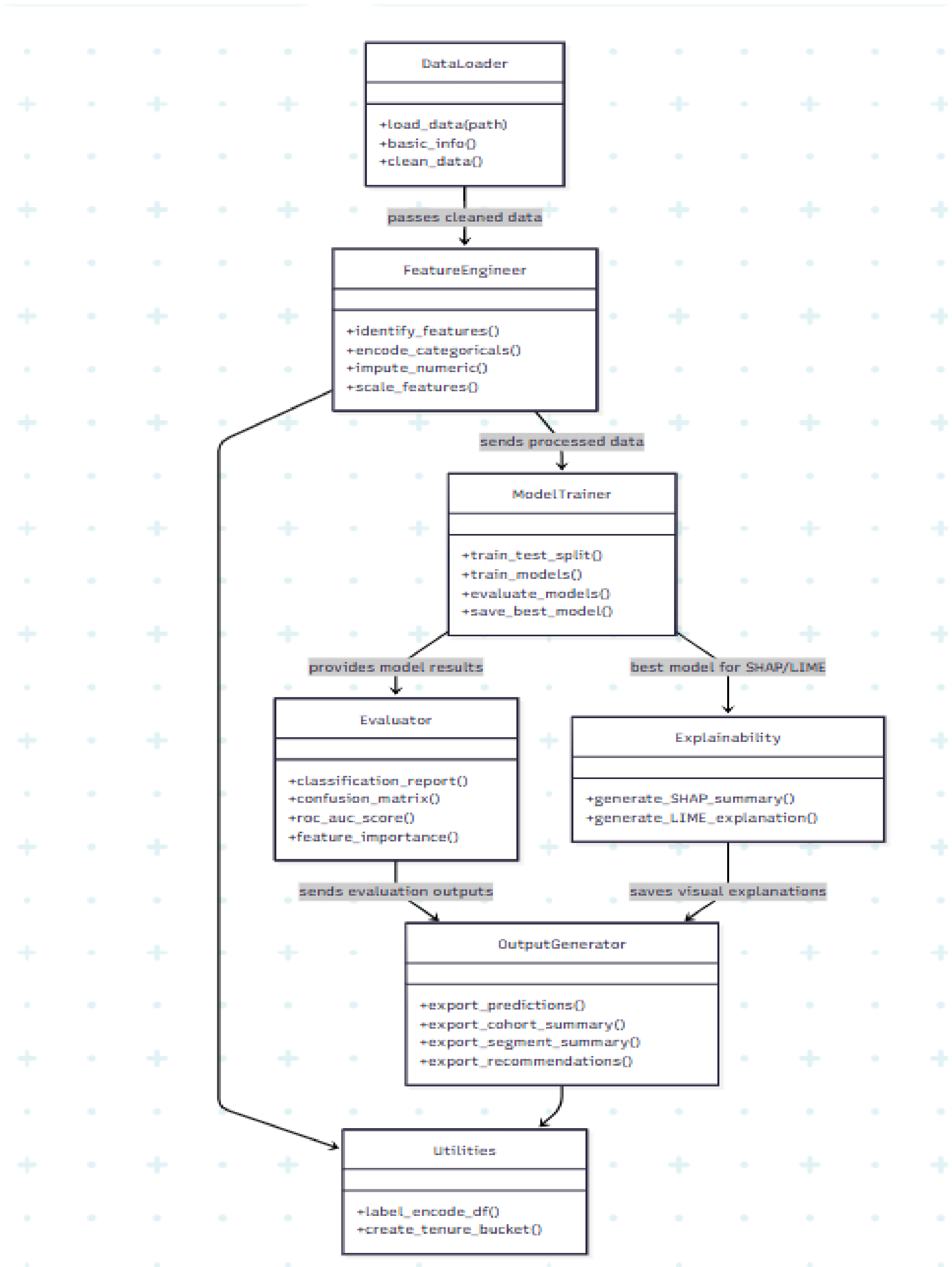



*Fig. 3. Detailed pipeline architecture of the proposed churn prediction system*

The overall methodological workflow of the proposed system includes the following stages:

- Dataset acquisition and exploratory data understanding.
- Data preprocessing and feature transformation.
- Feature engineering and tenure-based segmentation.
- Supervised machine learning model training and comparison.
- Model performance evaluation using multiple classification metrics.
- Explainable Artificial Intelligence analysis using SHAP and LIME.
- Integration of prediction outputs with Microsoft.
- Power BI dashboards for decision-support visualization

### 3.1 Dataset Description

The dataset used in this study is a publicly available telecom customer churn dataset containing **7,043 customer records and 21 service-related attributes**. The dataset includes demographic details, contract information, billing characteristics, service subscription features, and tenure duration of customers. The target variable *Churn* represents whether a customer discontinued telecom services or remained active.

Important predictive attributes include:

- Contract type
- Tenure duration
- Internet service type
- Payment method
- Monthly charges
- Total charges

- Technical support availability
- Online security subscription

These attributes provide valuable indicators for identifying customer behavior patterns related to churn prediction.

### 3.2 Data Preprocessing

Data preprocessing plays a critical role in improving model performance and ensuring dataset consistency before training machine learning algorithms. Several preprocessing steps were applied to transform raw telecom customer data into model-ready format.

Initially, missing values present in the TotalCharges attribute were handled using median imputation techniques to maintain dataset completeness. Categorical attributes such as contract type, payment method, and internet service type were converted into numerical representations using label encoding to enable compatibility with machine learning models. Numerical features including tenure duration, monthly charges, and total charges were standardized using feature scaling techniques to ensure balanced contribution across variables.

Finally, the dataset was divided into training and testing subsets using stratified sampling, where 75% of the data was used for training and 25% for testing while preserving the churn class distribution in both subsets.

### 3.3 Feature Engineering

Feature engineering techniques were applied to improve predictive capability of the classification models. Customers were grouped into tenure-based segments to identify early-stage churn risk patterns. Derived attributes such as tenure buckets improved interpretability and enabled better understanding of churn behavior across different subscription durations.

Additionally, preliminary feature importance analysis was performed during initial model training stages to identify key variables contributing significantly to churn prediction. Attributes such as contract type, tenure duration, monthly charges, and payment method consistently demonstrated strong predictive influence across classification models.

### 3.4 Machine Learning Model Selection

Customer churn prediction is formulated as a binary classification problem. Therefore, multiple supervised machine learning algorithms were evaluated to identify the most suitable prediction model.

The following classification models were implemented:

- Logistic Regression
- Random Forest
- XGBoost

Logistic Regression was used as a baseline classifier due to its interpretability and simplicity. XGBoost improved predictive capability through gradient boosting techniques that sequentially correct classification errors. Random Forest, an ensemble learning algorithm based on bootstrap aggregation, demonstrated the best performance among the evaluated models by capturing nonlinear relationships between telecom customer attributes and churn behavior.

Based on comparative evaluation using performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC, the **Random Forest classifier was selected as the final prediction model** due to its robustness, stability, and strong generalization capability.

### 3.5 Model Evaluation Metrics

To assess predictive performance of classification models, multiple evaluation metrics were used:

- Accuracy – measures overall classification correctness
- Precision – evaluates reliability of predicted churn customers
- Recall – measures effectiveness in identifying actual churn customers
- F1-score – balances precision and recall performance
- AUC-ROC – evaluates model discrimination capability between churn and non-churn classes

These evaluation metrics ensure both technical reliability and business applicability of the prediction model.

### 3.6 Explainable Artificial Intelligence Integration

Although ensemble learning techniques such as Random Forest provide strong predictive performance, interpretability remains essential for business decision-making. Therefore, explainable artificial intelligence techniques were integrated into the prediction framework.

SHAP (SHapley Additive Explanations) was applied to identify global feature importance across the dataset and determine the contribution of each attribute toward churn prediction. The analysis

revealed that contract type, tenure duration, monthly charges, and payment method significantly influence churn probability.
LIME (Local Interpretable Model-Agnostic Explanations) was used to generate instance-level explanations that describe why a specific customer is predicted to churn. These local explanations support personalized retention strategy planning and improve managerial trust in predictive analytics systems.

### 3.7 Power BI Dashboard Integration

To improve operational usability of churn prediction outputs, model predictions were integrated with Microsoft Power BI dashboards. Prediction probabilities generated by the Random Forest classifier were exported and visualized through interactive dashboards to analyze churn trends, customer segmentation patterns, and revenue distribution insights.

The Power BI dashboard enables decision-makers to:

- Identify high-risk customer groups
- Evaluate churn behavior across contract categories
- Monitor tenure-based churn variations
- Analyze revenue impact associated with customer attrition

This integration transforms predictive analytics outputs into actionable business intelligence for telecom retention planning.

## 4. RESULTS AND DISCUSSION

This section presents the experimental results obtained from multiple machine learning classification models implemented for telecom customer churn prediction. The performance of Logistic Regression, XGBoost, and Random Forest classifiers was evaluated using standard classification metrics including Accuracy, Precision, Recall, F1-score, and AUC-ROC. In addition to predictive evaluation, explainable artificial intelligence techniques and business intelligence dashboards were used to analyze churn drivers and business-level retention insights.

### 4.1 Model Performance Comparison

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2026, Volume 7 Issue 2
www.ijisea.org

To identify the most suitable classifier for churn prediction, three supervised learning models were implemented and evaluated. Logistic Regression served as a baseline classification model and provided moderate prediction accuracy with high interpretability. However, its performance was limited in capturing nonlinear relationships among telecom service attributes.
The XGBoost classifier improved classification performance using gradient boosting techniques that sequentially reduce prediction error. Although it demonstrated competitive predictive capability, its sensitivity to parameter tuning and dataset imbalance slightly affected overall generalization performance.
Among the evaluated models, the Random Forest classifier achieved the best balance between predictive accuracy, stability, and interpretability. Its ensemble learning structure effectively captured complex feature interactions between tenure duration, contract type, payment method, and service subscription attributes. Therefore, Random Forest was selected as the final churn prediction model for further analysis and deployment within the proposed retention intelligence framework.

The comparative performance of classification models is summarized below:

- Logistic Regression – Accuracy: 0.747, AUC: 0.843
- Random Forest – Accuracy: 0.766, AUC: 0.844
- XGBoost – Accuracy: 0.770, AUC: 0.804

Although XGBoost achieved slightly higher accuracy, Random Forest demonstrated better overall discrimination capability and model stability across evaluation metrics, making it more suitable for telecom churn prediction applications.

*Table1: Results*

| Model | Accuracy | Precision | Recall | F1 | AUC |
|---|---|---|---|---|---|
| Logistic Regression | 0.747 | 0.52 | 0.80 | 0.63 | 0.843 |
| Random Forest | 0.766 | 0.54 | 0.75 | 0.63 | 0.844 |
| XGBoost | 0.770 | 0.58 | 0.50 | 0.54 | 0.845 |

### 4.2 Random Forest Model Evaluation

The Random Forest classifier demonstrated strong classification capability in identifying both churn and non-churn customers. The confusion matrix analysis shows that the model correctly classified a high number of churners and retained customers with relatively fewer false negatives. Minimizing false negatives is particularly important in telecom churn prediction systems because these represent customers who are predicted as retained but actually discontinue services, directly contributing to revenue loss. The balanced classification performance achieved by the Random Forest model confirms its suitability for proactive retention analytics applications.

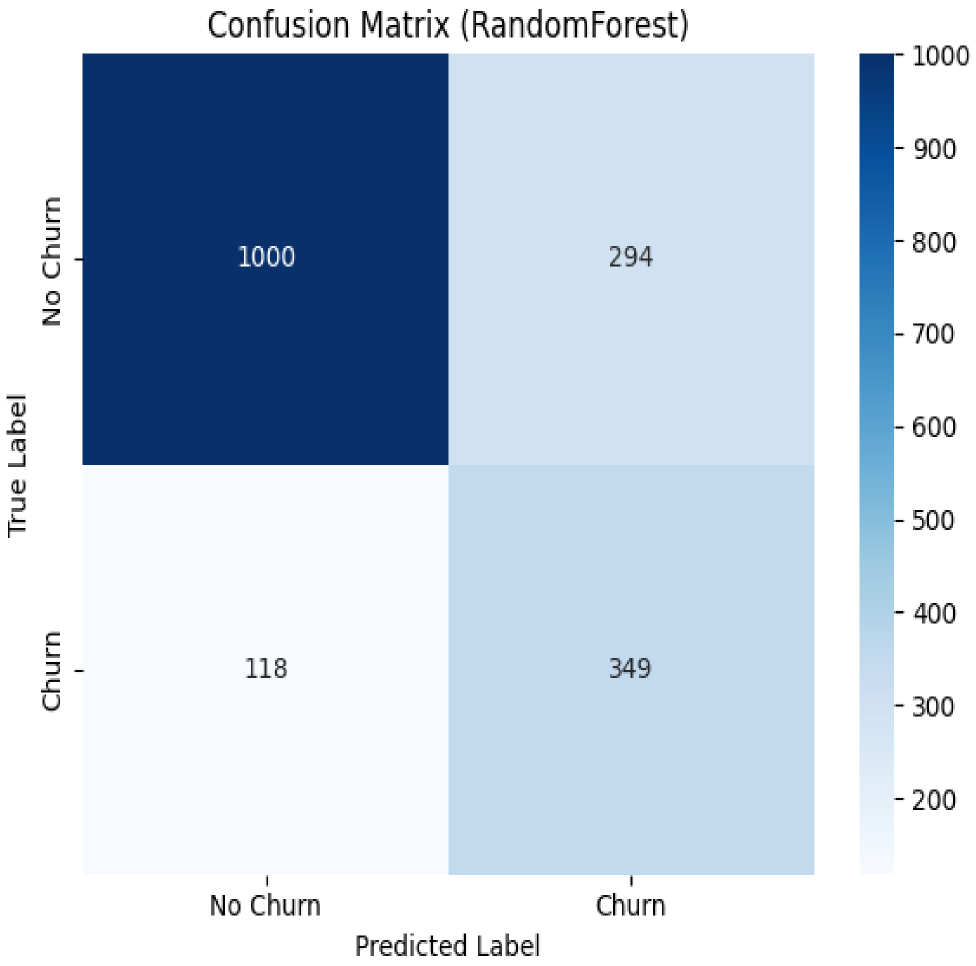



*Fig. 5 Confusion matrix of Random Forest churn classifier*

**4.3 ROC–AUC Performance Analysis**

The Receiver Operating Characteristic (ROC) curve was used to evaluate classification performance across multiple threshold values. The obtained ROC–AUC score of 0.844 indicates strong discriminative capability between churn and non-churn customer classes.

A higher ROC–AUC value demonstrates that the classifier effectively ranks churn customers with higher prediction probability compared to retained customers. This confirms that the proposed model performs reliably in identifying high-risk customers within telecom datasets.

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2026, Volume 7 Issue 2
www.ijisea.org

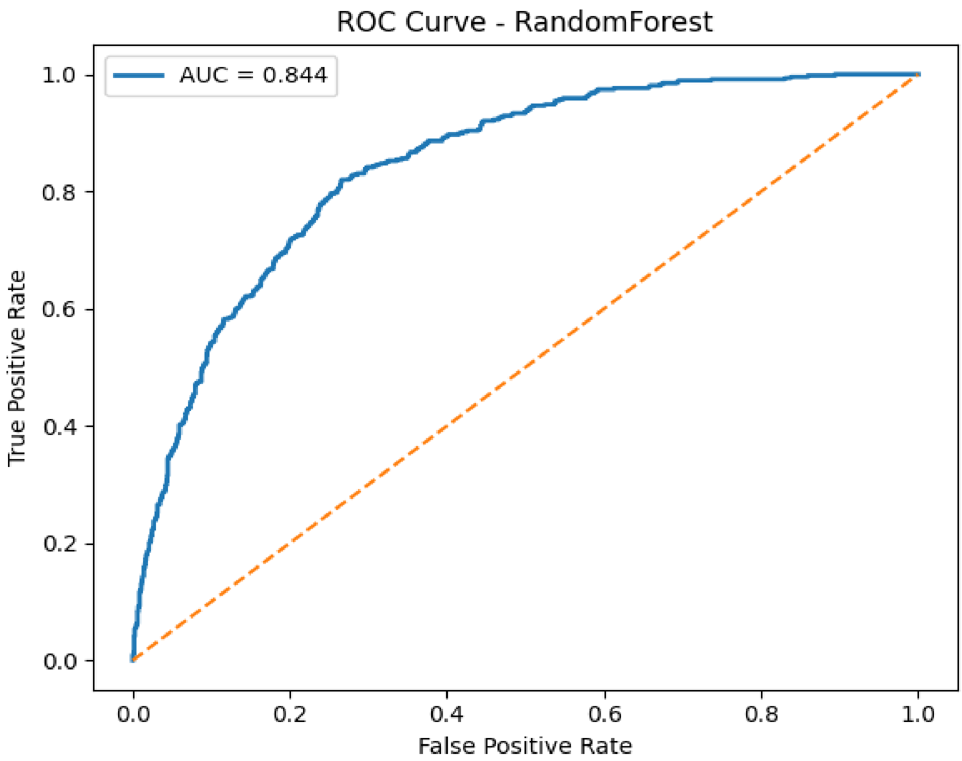



*Fig. 6 ROC curve showing classification performance of Random Forest model*

### 4.4 Feature Importance Analysis

Feature importance analysis was performed using the Random Forest classifier to identify the most influential attributes contributing to churn prediction. The most significant predictors identified include:

- Contract type
- Tenure duration
- Monthly charges
- Internet service category
- Payment method

Customers with month-to-month contracts showed higher churn probability compared to long-term contract subscribers. Similarly, customers with shorter tenure duration demonstrated increased churn risk, indicating that early-stage engagement plays an important role in retention

strategy planning. Higher monthly charges were also positively associated with churn behavior, suggesting pricing sensitivity among telecom subscribers.
These insights enable telecom service providers to identify high-risk customer segments and implement targeted retention strategies more effectively.

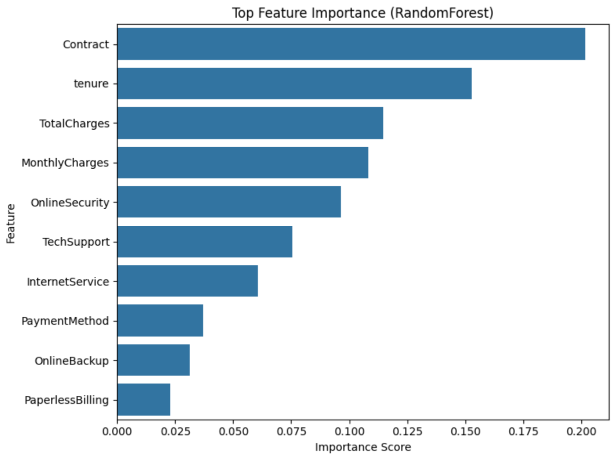



*Fig. 7 Feature importance ranking generated by Random Forest model*

**4.5 SHAP-Based Model Interpretation**

To improve interpretability of churn prediction results, SHAP explainability analysis was applied to identify both global and local feature contributions influencing model predictions.
The SHAP summary plot revealed that:

- Customers with shorter tenure duration exhibit higher churn probability
- Month-to-month contract customers represent the highest churn-risk segment
- Automatic payment methods reduce churn likelihood
- Higher monthly charges increase churn probability

The SHAP explanation framework enhances transparency of machine learning predictions and allows decision-makers to validate predictive insights against domain knowledge. These interpretability capabilities strengthen trust and support deployment of explainable predictive analytics solutions in telecom environments.

ISSN: 2582 - 6379
IJISEA Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2026, Volume 7 Issue 2
www.ijisea.org

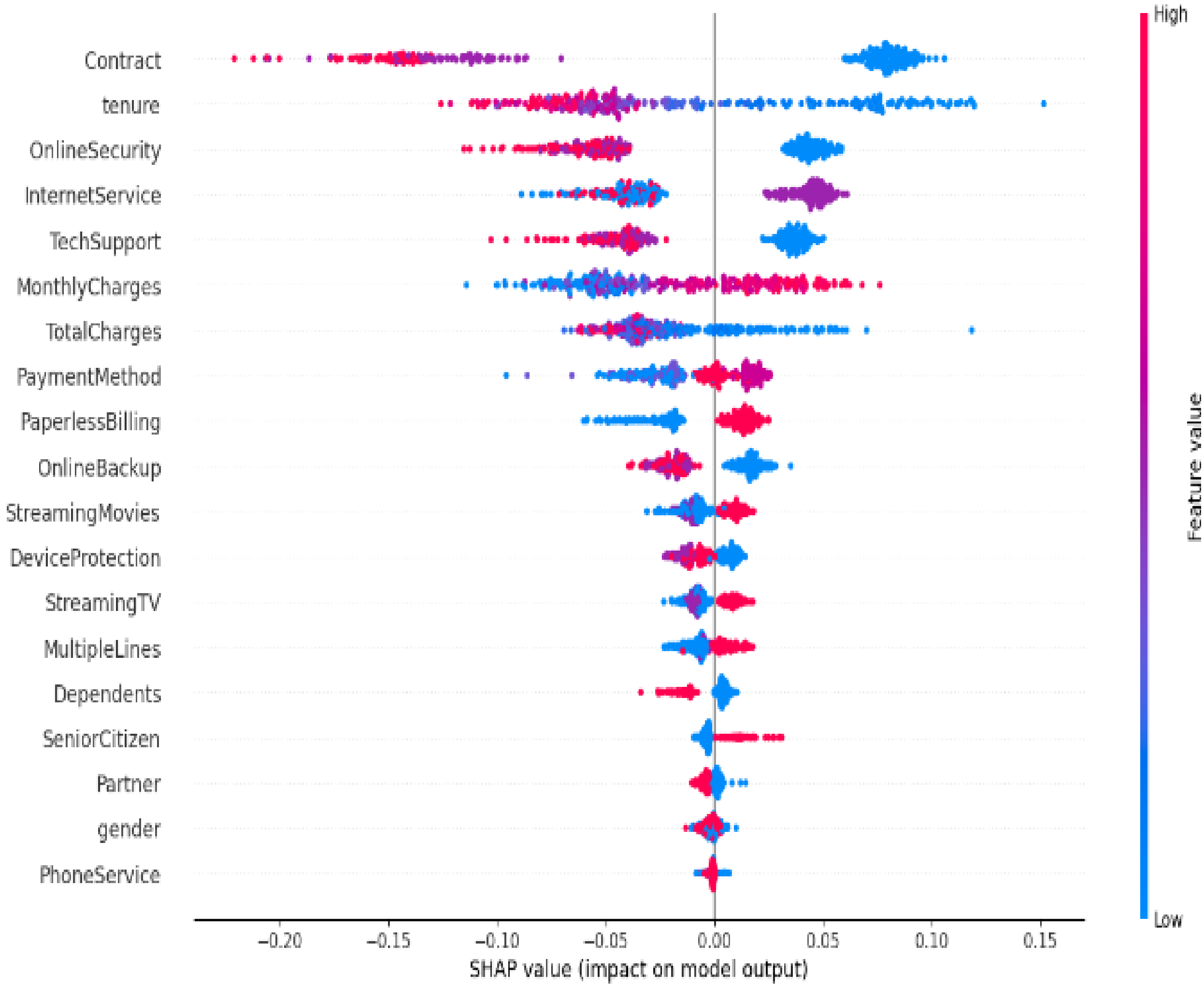



*Fig. 8 SHAP global explanation of churn prediction features*

### 4.6 Revenue Distribution Analysis

In addition to classification performance evaluation, revenue distribution analysis was performed to understand the financial impact of customer churn behavior.

The analysis revealed that a significant portion of total revenue is associated with customers who discontinue telecom services, highlighting the importance of early churn prediction for minimizing revenue loss. By linking churn predictions with billing attributes such as monthly charges and total charges, the proposed framework enables telecom organizations to prioritize retention strategies for high-value customers.

This business-oriented evaluation strengthens the practical applicability of the proposed churn prediction system.

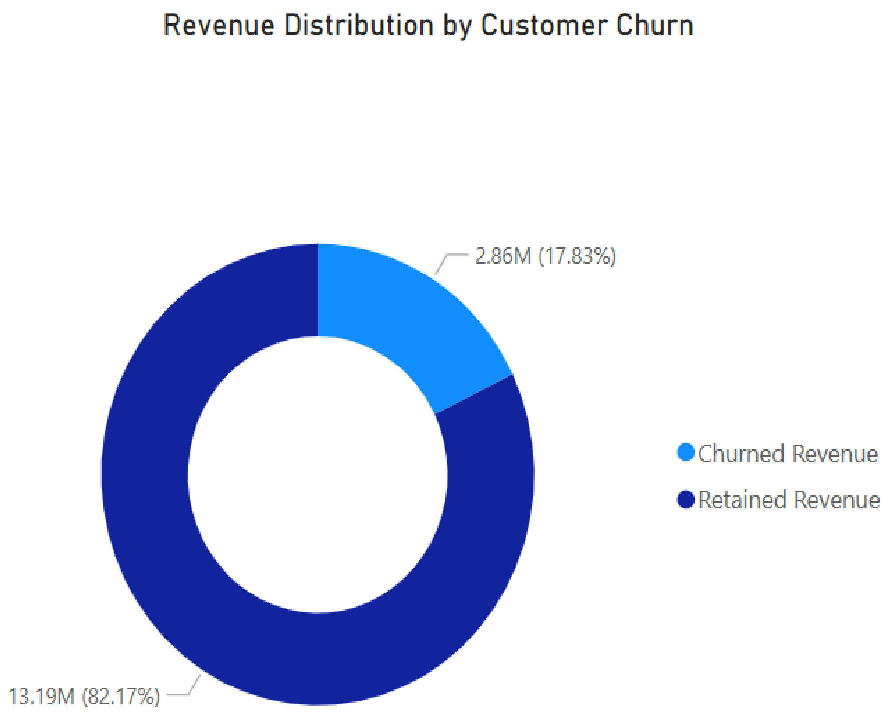


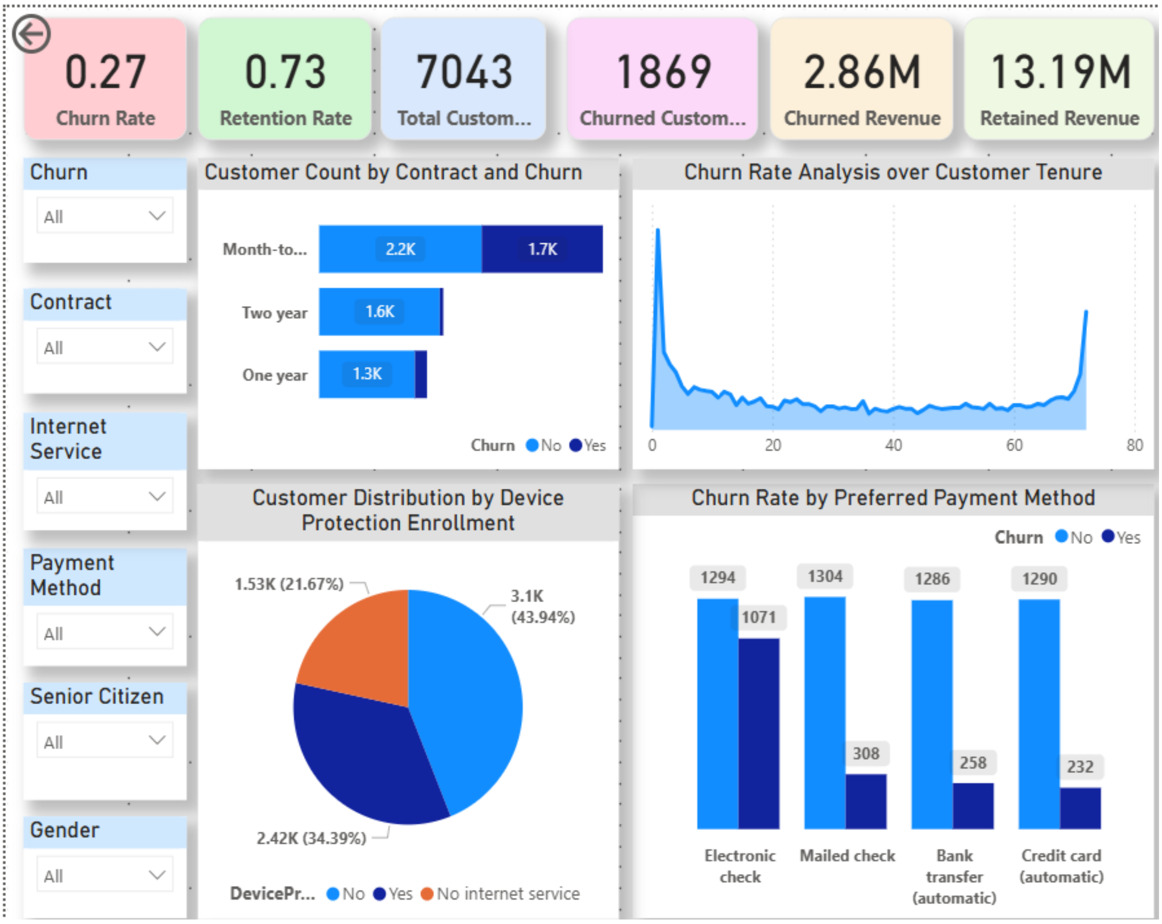

*Fig. 9 Revenue contribution comparison between churned and retained customers*

### 4.7 Power BI Dashboard Insights

The integration of machine learning predictions with Power BI dashboards enabled visualization of churn trends and customer segmentation patterns through interactive business intelligence components.
Key insights derived from dashboard analysis include:

- Customers with month-to-month contracts represent the highest churn-risk segment
- Electronic check payment users show increased churn probability
- Churn probability decreases significantly with increasing customer tenure
- Early subscription-stage customers require targeted retention engagement strategies

The interactive filtering capabilities of the dashboard further improve managerial decision-making by allowing segmentation of churn insights across demographic and service-based attributes. This integration transforms predictive analytics outputs into actionable retention intelligence for telecom service providers.


0.27
Churn Rate
0.73
Retention Rate
7043
Total Custom...
1869
Churned Custom...
2.86M
Churned Revenue
13.19M
Retained Revenue
Churn
All
Contract
All
Internet Service
All
Payment Method
All
Senior Citizen
All
Gender
All
Customer Count by Contract and Churn
Month-to...
2.2K
1.7K
Two year
1.6K
One year
1.3K
Churn No Yes
Churn Rate Analysis over Customer Tenure
0
20
40
60
80
Customer Distribution by Device Protection Enrollment
1.53K (21.67%)
3.1K (43.94%)
2.42K (34.39%)
DevicePr... No Yes No internet service
Churn Rate by Preferred Payment Method
Churn No Yes
1294
1071
1304
308
1286
258
1290
232
Electronic check
Mailed check
Bank transfer (automatic)
Credit card (automatic)

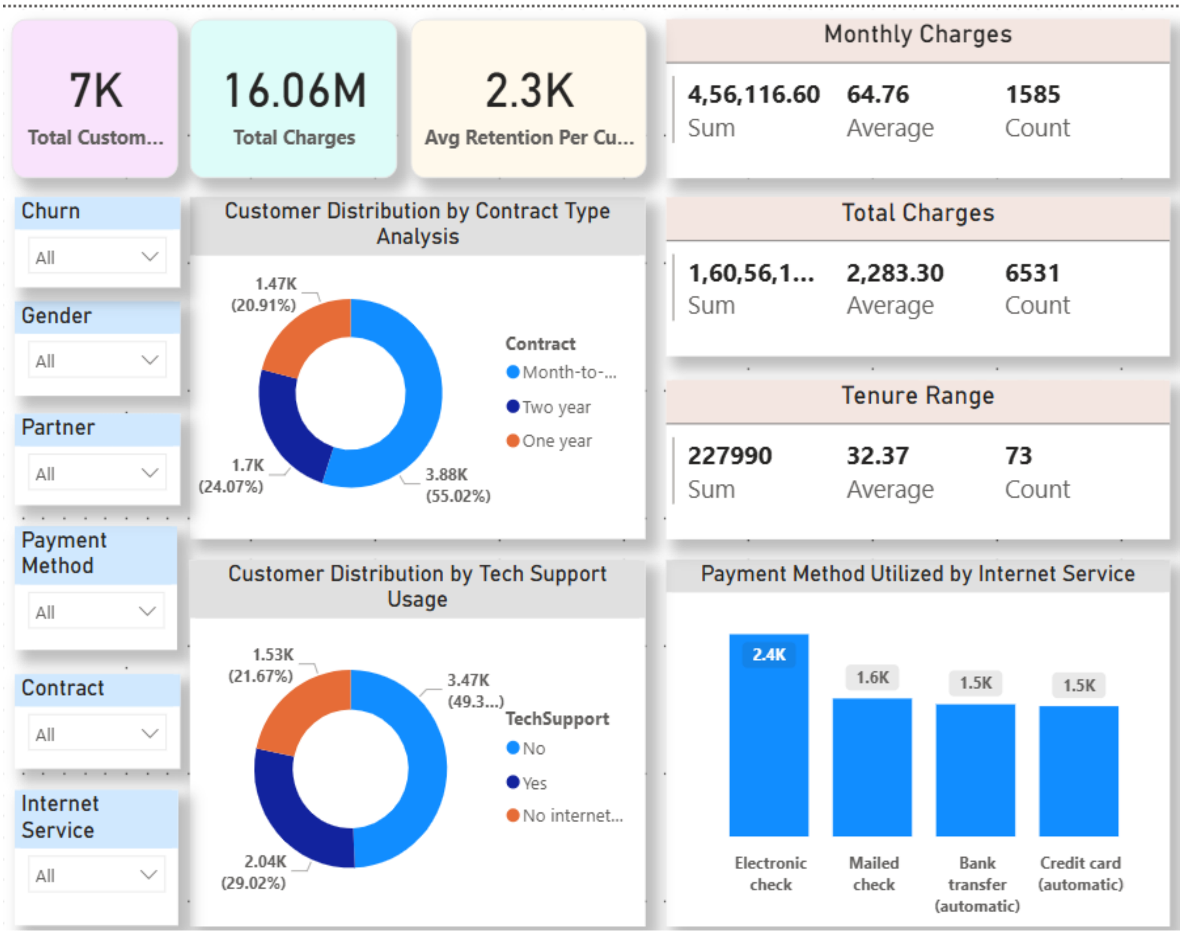



*Fig. 10 Interactive Power BI dashboard for churn analytics visualization*

## 5. CONCLUSION

This study proposed an explainable retention intelligence framework for telecom customer attrition prediction using ensemble machine learning techniques and interactive business intelligence visualization. Logistic Regression, Random Forest, and XGBoost models were evaluated using a publicly available Telco Customer Churn dataset obtained from Kaggle, where the Random Forest classifier achieved strong predictive performance with an accuracy of 0.7660 and ROC–AUC score of 0.8440.

The results demonstrate that the proposed framework can effectively identify high-risk churn customers using interpretable machine learning predictions supported by SHAP and LIME explanation techniques. Feature importance analysis revealed that contract type, tenure duration, monthly charges, and payment method are the most influential drivers of customer attrition behavior. By enabling early identification of churn-risk segments through predictive analytics and dashboard-based visualization, the proposed framework supports telecom service providers in implementing targeted retention strategies that can reduce potential customer loss and improve long-term revenue stability.

Future research can extend the proposed retention intelligence framework in several important directions:

- Incorporating deep learning architectures such as Artificial Neural Networks and Long Short-Term Memory (LSTM) models to capture complex temporal customer behavior patterns.
- Developing real-time churn prediction pipelines using streaming customer interaction data for continuous retention monitoring.
- Integrating customer sentiment analysis from service feedback records, call-center logs, and social media interactions using natural language processing techniques.
- Designing automated retention recommendation systems that trigger personalized intervention strategies when churn-risk probability exceeds predefined thresholds.
- Extending the proposed framework using graph-based learning techniques such as Graph Neural Networks to analyze customer relationship influence patterns affecting churn behavior.
- Addressing current dataset limitations by incorporating longitudinal telecom usage data for improving prediction robustness across time-dependent behavioral variations.

## REFERENCES

[1] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Morgan Kaufmann, 2011.

[2] Géron A, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd ed. O'Reilly Media, 2019.

[3] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[4] J. R. Quinlan, "Induction of decision trees," Machine Learning, vol. 1, no. 1, pp. 81–106, 1986.

[5] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[6] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NeurIPS), 2017.

[7] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016.

[8] W. Verbeke, D. Martens, C. Mues, and B. Baesens, "Building comprehensible customer churn prediction models with advanced rule induction techniques," Expert Systems with Applications, vol. 38, no. 3, pp. 2354–2364, 2011.

[9] J. Hadden, A. Tiwari, R. Roy, and D. Ruta, "Computer-assisted customer churn management: State-of-the-art and future trends," Computers & Operations Research, vol. 34, no. 10, pp. 2902–2917, 2007.

[10] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," Decision Support Systems, vol. 50, no. 3, pp. 559–569, 2011.

[11] Microsoft Corporation, "Power BI documentation," Microsoft Official Documentation, 2023. [Online]. Available: https://learn.microsoft.com/power-bi

[12] F. Chollet, Deep Learning with Python. Manning Publications, 2018.

[13] F. Provost and T. Fawcett, Data Science for Business. O'Reilly Media, 2013.

[14] H. Witten, E. Frank, and M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, 2016.

[15] Kaggle, "Telco Customer Dataset." [Online]. Available: https://www.kaggle.com/code/bhartiprasad17/customer-churn-prediction

[16] S. Amin, A. Anwar, A. Adnan, M. Nawaz, K. Alawfi, A. Hussain, and K. Huang, "Customer churn prediction in telecommunication industry using data certainty," Journal of Business Research, vol. 94, pp. 290–301, 2019.

[17] H. Idris, A. Khan, and Y. S. Lee, "Intelligent churn prediction in telecom: Employing mRMR feature selection and RotBoost-based ensemble classification," Applied Intelligence, vol. 39, no. 3, pp. 659–672, 2013.