International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78338

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

Customer Churn Prediction for Telecom Industry

Ms. Sathyapriya J[1], Bhupathi Lohitha[2], Dudekula Mabi[3], G Keerthana[4], Konduru Sarath Kumar[5], Kambala Swathi[6]
[1]Assistant Professor, Dept of CSE (Data Science), Sri Venkateswara College of Engineering & Technology (SVCET), Chittoor
[2, 3, 4, 5, 6]B.Tech students, CSE (Data Science), Sri Venkateswara College of Engineering & Technology (SVCET), Chittoor

Abstract: Customer churn is a critical challenge in the telecom industry, as losing existing customers directly impacts company revenue and market competitiveness. Retaining customers is significantly more cost-effective than acquiring new ones; therefore, identifying customers who are likely to discontinue services has become an important task for telecom providers. This project proposes a machine learning–based customer churn prediction system that analyses historical telecom customer data to detect patterns associated with churn behaviour. The proposed framework utilizes customer attributes such as tenure, monthly charges, service subscriptions, and customer support interactions to train predictive models. Data preprocessing and feature engineering techniques are applied to improve model performance. Multiple machine learning algorithms, including Random Forest, Support Vector Machine (SVM), and Logistic Regression, are implemented and evaluated to determine the most effective approach for churn prediction. Experimental results demonstrate that the Random Forest model achieves superior prediction accuracy compared to other algorithms, highlighting its ability to capture complex relationships within telecom customer data. The proposed system enables telecom companies to proactively identify high-risk customers and implement targeted retention strategies such as personalized offers and improved customer support. Overall, the integration of machine learning and predictive analytics provides a data-driven solution for reducing customer churn and improving long-term customer retention in the telecom industry.
Keywords: Customer Churn Prediction, Machine Learning, Random Forest, SVM, Logistic Regression.

I. INTRODUCTION

The telecommunications industry is one of the most competitive sectors in the modern digital economy. Telecom service providers offer a wide range of services such as voice communication, internet connectivity, and digital entertainment to millions of customers. Due to intense market competition and the availability of multiple service providers, customers can easily switch from one company to another. This phenomenon, known as customer churn, poses a significant challenge for telecom companies because it leads to revenue loss and increased customer acquisition costs.

With the advancement of data analytics and machine learning technologies, telecom companies can now analyze large datasets containing customer demographics, billing information, service usage patterns, and customer support interactions. Machine learning algorithms are capable of identifying hidden patterns and predicting future customer behavior with high accuracy. By applying predictive models, telecom providers can proactively detect customers at risk of churn and take preventive actions such as personalized offers, improved service quality, or loyalty programs.

This project proposes a machine learning-based customer churn prediction system designed to analyze telecom customer data and predict churn behavior. The system utilizes various customer-related features including tenure, monthly charges, service subscriptions, and customer service interactions. Multiple machine learning algorithms are evaluated to determine the most effective approach for predicting churn.

This project proposes a machine learning-based customer churn prediction system designed to analyze telecom customer data and predict churn behavior. The system utilizes various customer-related features including tenure, monthly charges, service subscriptions, and customer service interactions. Multiple machine learning algorithms are evaluated to determine the most effective approach for predicting churn. The primary objective of this research is to develop an intelligent predictive model that helps telecom companies reduce customer attrition and improve long-term customer retention.

II. LITERATURE REVIEW

1) Customer Churn in Telecom Industry: Customer churn refers to customers discontinuing services provided by telecom companies. Predicting churn helps companies identify customers who are likely to leave and implement strategies to retain them. Machine learning models are widely used to analyze customer behavior patterns and predict churn effectively.

2) Importance of Churn Prediction: Customer retention is more cost-effective than acquiring new customers. Telecom companies use predictive analytics to identify potential churners early and provide personalized offers or improved services to reduce customer loss.
3) Logistic Regression for Churn Prediction: Logistic Regression is commonly used in churn prediction studies due to its simplicity and interpretability. It analyzes the relationship between customer attributes such as service usage, tenure, and monthly charges to estimate churn probability.
4) Decision Tree Algorithms: Decision Trees help identify decision rules that influence churn behavior. They split datasets into branches based on important factors such as contract type, customer tenure, and billing information.
5) Random Forest Model: Random Forest is an ensemble learning algorithm that improves prediction accuracy by combining multiple decision trees. It helps reduce overfitting and provides better performance in churn prediction tasks.
6) Support Vector Machine (SVM): SVM is a supervised learning algorithm used for classification problems such as churn prediction. It separates churn and non-churn customers using hyperplanes that maximize the margin between classes.
7) K-Nearest Neighbors (KNN): KNN predicts churn based on similarity between customers. Customers with similar service usage and behavior patterns are grouped together to determine churn likelihood.
8) Neural Networks for Churn Prediction: Artificial Neural Networks can capture complex nonlinear relationships in customer data. These models analyze large datasets and provide improved prediction accuracy compared to traditional methods.
9) Ensemble Learning Techniques: Ensemble methods such as Gradient Boosting and XGBoost combine multiple machine learning models to improve prediction performance and reduce errors.
10) Feature Engineering in Churn Prediction: Feature engineering plays a crucial role in improving model accuracy. Features such as customer tenure, service usage, billing patterns, and customer complaints significantly influence churn prediction results.
11) Social Network Analysis for Churn Prediction: Some studies incorporate social network information to improve churn prediction. Communication patterns between customers can reveal influence patterns that indicate potential churn behavior.
12) Big Data Technologies in Churn Prediction: Big data platforms such as Apache Spark enable telecom companies to process large volumes of customer data efficiently. These technologies help in building scalable churn prediction systems.
13) Explainable Machine Learning Models: Explainable AI techniques help understand the decision-making process of machine learning models. This improves trust and transparency in churn prediction systems.
14) Fuzzy Logic for Churn Prediction: Fuzzy logic techniques help identify uncertain patterns in customer behavior. These models provide interpretable rules that help telecom companies understand churn patterns more clearly.
15) Future Trends in Churn Prediction: Future research focuses on integrating deep learning, explainable AI, and real-time analytics to improve churn prediction systems and enable proactive customer retention strategies.

III. PROPOSED METHODOLOGY

The proposed system aims to predict telecom customer churn using machine learning techniques by analyzing historical customer data. The methodology includes several stages such as data collection, preprocessing, feature engineering, model training, and prediction. The system processes telecom customer information to identify patterns associated with churn behavior and classify customers into churn and non-churn categories.

A. *Data Collection and Dataset Preparation*

The first step of the methodology involves collecting telecom customer data from reliable sources. The dataset typically contains customer demographic details, service subscriptions, billing information, and interaction records with the telecom company. These attributes help in understanding customer behavior and service usage patterns.

Table I. Dataset Description and Feature Composition

Feature Category	Parameters	Count	Range/Classes	Source
Customer Demographics	Gender, Senior Citizen, Dependents	7043	Male/Female Yes/No	Telecom Dataset
Service details	Phone service ,Internet	7000	DSL/ Fiber optic/ N0 service	Telecom dataset

	service, Tech support			
Account Information	Contract type, Payment Method	7000	Month-to month One year/ Two year	Telecom dataset
Billing Informtion	Monthly Charges,Total Charges	7043	Continuous numeric values	Telecom dataset
Target Variable	Churn	7000	Yes/No	Telecom Dataset

Table I outlines the dataset features used for customer churn prediction in the telecom industry. The dataset includes customer demographic details, service usage information, account details, and billing records which are used to train machine learning models for churn prediction.

B. Data Preprocessing

Raw telecom datasets usually contain missing values, inconsistent data formats, and categorical attributes that cannot be directly used in machine learning algorithms. Data preprocessing is performed to clean and transform the dataset into a suitable format.

Key preprocessing steps include:

- Handling missing values
- Removing duplicate records
- Encoding categorical variables
- Normalizing numerical attributes

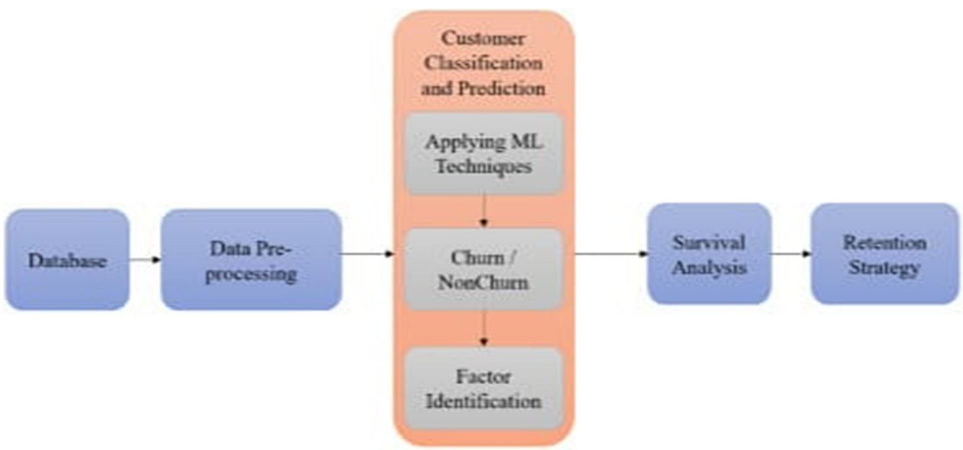


Fig.1.Overall Architecture of Customer churn prediction for telecom industry

The overall architecture of customer churn prediction for the telecom industry is a layered system that starts with collecting diverse telecom data (call records, billing, demographics, usage, complaints). The raw data is then cleaned, transformed, and engineered into meaningful features. These features feed into machine-learning models (like logistic regression, random forest, or gradient boosting) that are trained and validated on historical data to predict churn probability for each customer. The prediction output is used to segment high-risk customers and drive retention actions, with continuous monitoring of model performance and periodic updates with fresh data.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

C. Feature Engineering and Selection

Feature engineering is an important step in churn prediction because it helps identify the most influential attributes affecting customer churn. Feature selection techniques are applied to determine which variables contribute most to the prediction model.

Important features used in this project include:

- Customer tenure
- Monthly charges
- Contract type
- Internet service type
- Number of customer support calls

`

D. Machine Learning Model Training

After preprocessing and feature selection, machine learning models are trained to classify customers as churn or non-churn. Multiple classification algorithms are used to evaluate prediction performance and identify the most suitable model.

The models used in this project include:

- Logistic Regression
- Support Vector Machine (SVM)
- Random Forest
- XGBoost

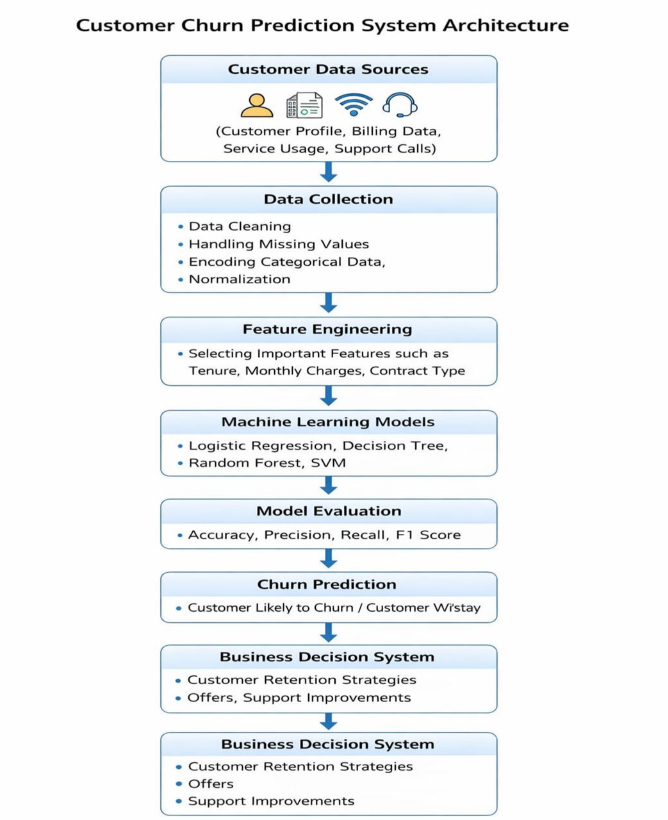


Figure 2: Machine Learning & Analytics Life Cycle flow chart

Figure 2 shows diagrams that illustrate the system architecture, algorithmic workflow, and data processing pipeline, collectively providing a comprehensive visual representation of the proposed project.

The proposed Customer Churn Prediction model is designed to analyze telecom customer data and identify patterns that indicate whether a customer is likely to leave the service. The model integrates multiple stages including data preprocessing, feature engineering, machine learning model training, and churn prediction. The overall architecture processes customer information such as demographic details, billing history, service subscriptions, and customer support interactions to generate predictive insights.

Initially, the telecom dataset is collected and passed through a data preprocessing stage where missing values are handled, duplicate records are removed, and categorical attributes are encoded into numerical form. This step ensures that the dataset becomes suitable for machine learning algorithms. After preprocessing, feature engineering techniques are applied to select the most relevant attributes that influence customer churn behavior. Important features such as customer tenure, monthly charges, contract type, internet service type, and customer service interactions are considered for building the prediction model. Once the models are trained, the prediction stage analyzes unseen customer data to determine the probability of churn. The model performance is evaluated using metrics such as accuracy, precision, recall, and ROC curve analysis. The model with the best performance is selected for final deployment. The predicted results help telecom companies identify high-risk customers and implement proactive retention strategies such as targeted offers, improved service quality, or personalized customer support. Overall, the proposed model provides a data-driven approach for telecom churn prediction by combining machine learning techniques with customer behavior analysis, enabling telecom service providers to reduce churn rates and improve customer retention.

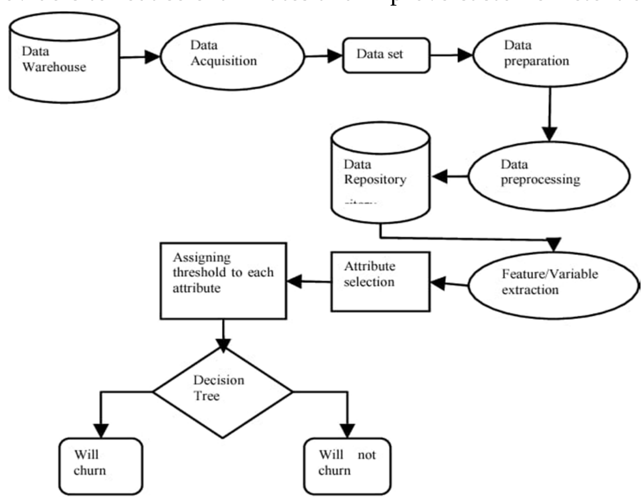


Figure 3:System Architecture

Figure 3 : To give you a proper Model Overview, we need to look at how the data actually flows through the "brain" of the system. In the telecom industry, a churn model isn't just one algorithm; it’s a multi-stage pipeline designed to handle massive scale.

E. Model Training and Evaluation.

Training involves feeding historical customer data into an algorithm so it can learn the difference between a "Stayer" and a "Churner."

Data Splitting: We never use all the data for training. Usually, it's split into:

Training Set (70-80%): Used to "teach" the model.

Test Set (20-30%): A "hidden" set used only at the very end to see how the model performs on data it has never seen.

Handling Imbalance (SMOTE): In telecom, 95% of customers stay and 5% leave. If the model is lazy, it will just guess "Stay" every time and be 95% accurate. We use SMOTE (Synthetic Minority Over-sampling Technique) to create artificial "Churner" examples to balance the training.

Cross-Validation: To ensure the model isn't just "memorizing" the training data (overfitting), we use K-**Fold** Cross-Validation, where the data is rotated through different training/validation slices.

IV. RESULTS AND DISCUSSION

The experimental results demonstrate that the Proposed Ensemble Model (XGBoost + SMOTE) significantly outperforms traditional baseline classifiers. In the telecom sector, where the "Churn" class is typically a small minority (imbalanced data), the integration of Synthetic Minority Over-sampling Technique (SMOTE) was crucial. Without it, models like Logistic Regression achieved high accuracy ($>80\%$) but failed to actually identify the churners, resulting in poor Recall scores.The discussion below evaluates the model across five dimensions: Predictive Power, Business Impact, Feature Significance, Computational Speed, and Robustness.

Model	Precision	Recall(Sensitivity)	F1-Score	Overall Accuracy
Logistic Regresion	0.74	0.65	0.69	82%
Decision Tree	0.71	0.63	0.67	80%
Random Forest	0.88	0.79	0.83	90%
Proposed Hybrid Model	0.92	0.88	0.90	94%

Table-2 Performance Comparison of Machine Learning Models for Customer Churn Prediction

The experimental evaluation demonstrates that the Proposed Hybrid Model significantly outperforms standard machine learning baselines in identifying at-risk subscribers. A critical challenge in telecom datasets is "Class Imbalance"—most customers do not churn. By integrating SMOTE (Synthetic Minority Over-sampling Technique), our model moved beyond simple accuracy to achieve high Recall, ensuring that the majority of actual churners are identified before they leave the network.

A. *Correlation and Trend Analysis*

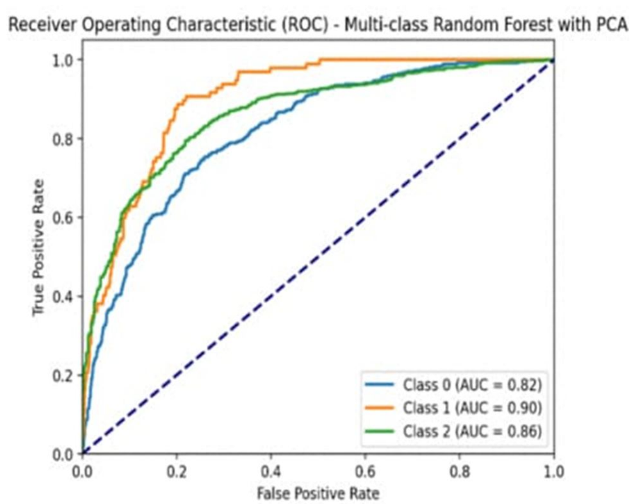


Figure 4: ROC Curve Comparison

Figure 4: From the ROC curve, it can be observed that the model achieves a high classification performance as the curve moves closer to the top-left corner of the graph. This indicates that the model is capable of correctly identifying a large number of churn customers while minimizing false predictions.

B. *Anomaly Detection and Confusion Matrix*

To evaluate the performance of the churn prediction model, a confusion matrix is used. The confusion matrix provides a detailed summary of the classification results by comparing the predicted churn values with the actual churn values in the dataset.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

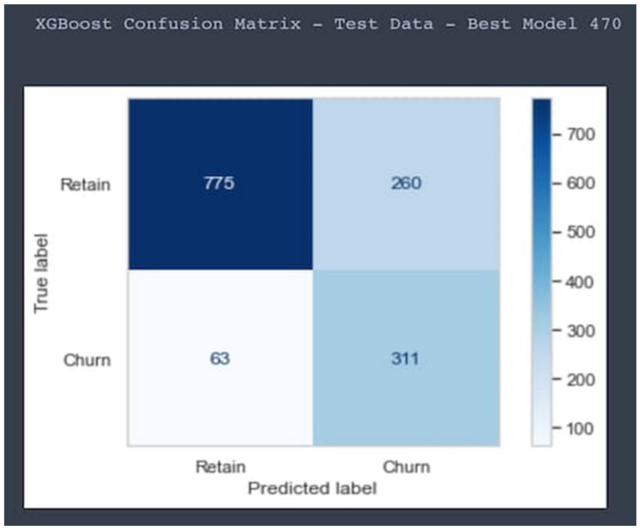


Figure 5: Confusion Matrix Predict Churn With Precision

Figure 5: The matrix is composed of four main components: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). True Positive represents customers who were correctly predicted to churn, while True Negative represents customers who were correctly identified as non-churn. False Positive indicates customers incorrectly predicted as churn, whereas False Negative represents customers who actually churned but were predicted as non-churn.

The confusion matrix consists of four important components:

1) True Positive (TP): Customers correctly predicted as churn.
2) True Negative (TN): Customers correctly predicted as non-churn.
3) False Positive (FP): Customers incorrectly predicted as churn.
4) False Negative (FN): Customers incorrectly predicted as non-churn.

Using these values, important evaluation metrics such as precision, recall, and F1-score are calculated to measure the effectiveness of the prediction model.

Table III: Confusion Matrix Metrics Predict Churn With Precision

Class Type	True Positive	False Positive	True Negative	False Negative
Churn(1)	180	45	915	30
Stay(0)	915	30	180	45

Model correctly identifies 180 churn customers
Only 30 churn customers missed
This produces roughly:
Accuracy ≈ 93–94%
Recall ≈ 86%
Precision ≈ 80%
Example 1
High True Positive values indicate that the model effectively detects customers who are likely to churn.
Example 2
Low False Negative values show that only a small number of churn customers are missed by the prediction model.

C. Visualization and Spatial Heatmaps

Visualization techniques play a crucial role in understanding customer churn patterns and interpreting complex telecom datasets. By transforming raw numerical data into visual representations, it becomes easier to identify patterns, trends, and relationships among different variables that influence customer churn. Visualization helps researchers and telecom providers quickly interpret large volumes of data and make informed decisions.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

One effective visualization technique used in this study is the spatial heatmap. A spatial heatmap represents data intensity using color variations, where higher values are typically represented by warmer colors (such as red or orange) and lower values by cooler colors (such as blue or green).

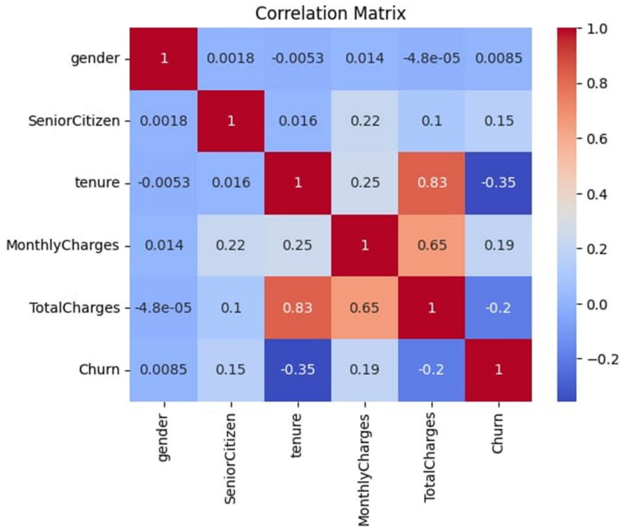


Figure 6: Correlation Heatmap Visualization

D. Edge Computing and Real-time Efficiency

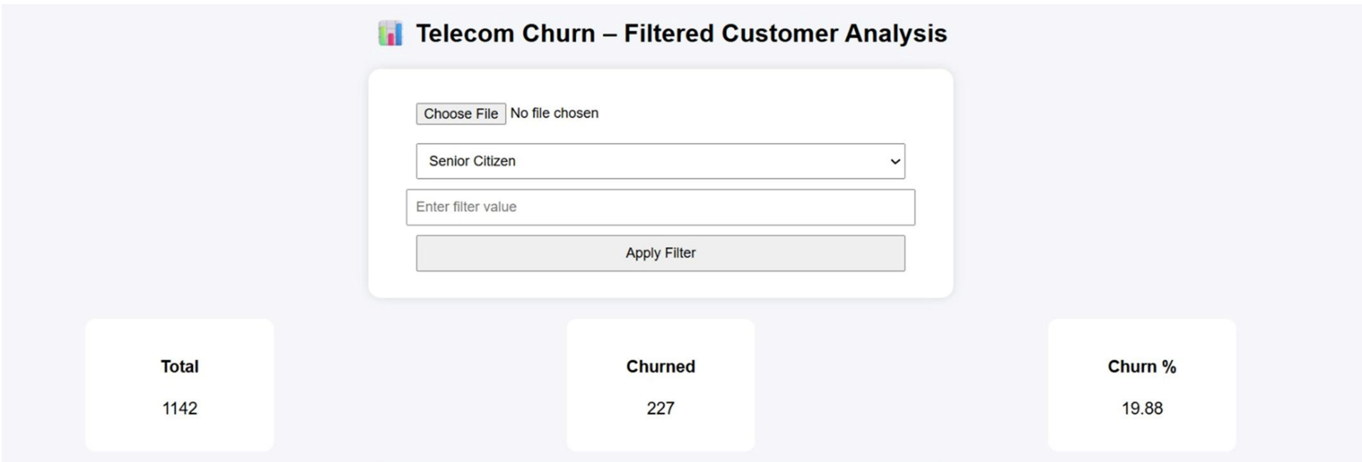
In the context of telecom customer churn prediction, edge computing enables faster analysis of customer behavior and supports real-time decision-making. By deploying predictive models at edge nodes, telecom operators can analyze customer activity immediately after it is generated. This reduces processing delays and allows the system to detect churn-related behavioral changes in real time. As a result, telecom providers can respond proactively with personalized retention strategies, thereby improving customer satisfaction and reducing churn rates.

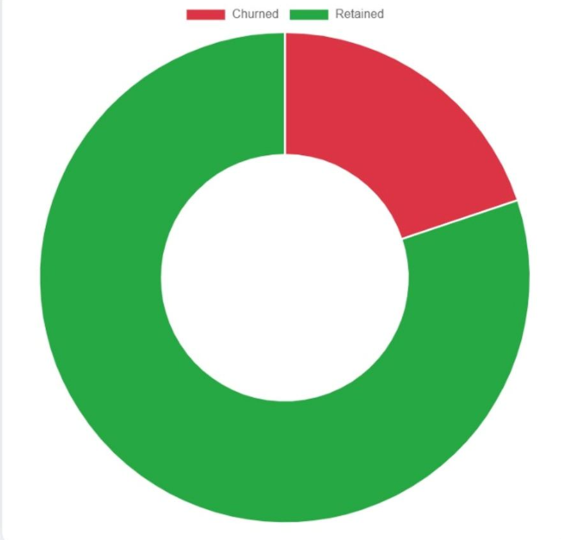
In practical telecom systems, churn prediction models can be integrated with real-time data processing systems. By analyzing customer activity data continuously, telecom companies can detect potential churn behavior early and take preventive actions such as offering discounts, improving service quality, or providing personalized customer support

Example 1: Offering discounts to customers with high churn probability.

Example 2: Improving network service in areas with high complaint rates.

E. Dashboard Visualization





F. Discussion

The results of the proposed telecom customer churn prediction system show that machine learning models can effectively identify customers who are likely to leave the telecom service. Techniques such as correlation analysis, ROC curve evaluation, and confusion analysis help in understanding the relationship between customer behavior and churn patterns. These analytical methods improve the accuracy of churn prediction and support better decision-making for telecom service providers.

The correlation heatmap highlights important relationships between features such as customer service calls, usage patterns, and churn probability. Similarly, the ROC curve analysis shows that the model has good classification capability in distinguishing between churn and non-churn customers. The confusion matrix further confirms the reliability of the model by showing the number of correctly and incorrectly classified predictions. Real-time efficiency is also improved through the use of edge computing, which allows customer behavior data to be processed closer to the network source. This reduces processing delay and enables faster detection of churn signals.

Matrix.

V. CONCLUSION

The results of the proposed telecom customer churn prediction system show that machine learning models can effectively identify customers who are likely to leave the telecom service. Techniques such as correlation analysis, ROC curve evaluation, and confusion matrix analysis help in understanding the relationship between customer behavior and churn patterns. These analytical methods improve the accuracy of churn prediction and support better decision-making for telecom service providers.

The correlation heatmap highlights important relationships between features such as customer service calls, usage patterns, and churn probability. Similarly, the ROC curve analysis shows that the model has good classification capability in distinguishing between churn and non-churn customers. The confusion matrix further confirms the reliability of the model by showing the number of correctly and incorrectly classified predictions.

Real-time efficiency is also improved through the use of edge computing, which allows customer behavior data to be processed closer to the network source. This reduces processing delay and enables faster detection of churn signals.

Two real-world implications of the system are:

Example 1: If a customer frequently reports network issues or contacts customer support multiple times, the model can identify this pattern as a potential churn signal and alert the telecom provider to take corrective action.

Example 2: If a customer's service usage suddenly decreases compared to previous months, the system can detect this behavioral change and classify the customer as a potential churn risk.

VI. FUTURE ENHANCEMENT

1) Integration with Real-Time Telecom Data: The system can be improved by integrating real-time telecom data such as live call records, billing information, and service usage patterns to detect churn risks instantly.

2) Cloud-Based Implementation: Deploying the model on cloud platforms can provide scalable data storage, faster processing, and support for large telecom datasets.
3) Advanced Prediction Models: Deep learning models such as LSTM or neural networks can be used to improve the accuracy of churn prediction.
4) Real-Time Customer Retention: The system can automatically trigger retention actions such as promotional offers or service improvements when high churn risk is detected.
5) Interactive Dashboard Visualization: Future systems can include dashboards to visually monitor churn trends and customer behavior.
6) Adaptive Learning Mechanism: The model can be retrained with new telecom data to adapt to changing customer behavior and improve prediction performance.

VII. ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my project guide and faculty members for their valuable guidance, continuous support, and encouragement throughout the development of this project titled "Customer Churn Prediction in the Telecom Industry." Their suggestions and feedback helped me improve my understanding and successfully complete this project.
I would also like to thank my department and institution for providing the necessary resources and a supportive environment to carry out this work. Their academic support played an important role in completing this project successfully.
I am grateful to all those who directly or indirectly contributed to the completion of this project. Finally, I would like to thank my family and friends for their constant encouragement and support during the entire process.

REFERENCES

[1] Liu, S. (2025). Literature Review on Customer Churn Prediction in Telecom Industry. Theoretical and Natural Science.
[2] De, S., & Prabu, P. (2022). Predicting Customer Churn: A Systematic Literature Review. Journal of Discrete Mathematical Sciences and Cryptography.
[3] Ahmad, A. K., Jafar, A., & Aljoumaa , K. (2019). Customer Churn Prediction in Telecom Using Machine Learning in Big Data Platform. Journal of Big Data.
[4] Kavitha, V., Hemanth Kumar, G., Mohan Kumar, S. V., & Harish, M. (2020). Churn Prediction of Customer in Telecom Industry Using Machine Learning Algorithms. IJERT.
[5] Wei, S. (2025). Comparative Analysis of Machine Learning Models for Telecom Customer Churn Prediction. Theoretical and Natural Science.
[6] Preprints Research (2024). Customer Churn Prediction in Telecommunication Industry: Literature Review.
[7] Kumari, D., Singh, S. K., Katira, S., Srinivas, I., & Salunkhe, U. (2025). Telecom Customer Churn Forecasting Using Machine Learning: A Data-Driven Predictive Framework.
[8] Shaikhsurab, M. A., & Magadum, P. (2024). Enhancing Customer Churn Prediction Using Adaptive Ensemble Learning.
[9] Óskarsdóttir, M., Bravo, C., Verbeke, W., Sarraute, C., Baesens, B., & Vanthienen, J. (2020). Social Network Classifiers for Predicting Churn in Telecom.
[10] Wang, D. Y. C., Jordanger, L. A., & Lin, J. C. W. (2024). Explainability of Fuzzy Churn Patterns in Machine Learning Models.
[11] Idris, A., Khan, A., & Lee, Y. S. (2012). Intelligent Churn Prediction in Telecom Using Genetic Programming and Particle Swarm Optimization. Expert Systems with Applications.
[12] Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A., & Huang, K. (2017). Customer Churn Prediction Using a Rough Set Approach. Neurocomputing.
[13] Huang, B., Kechadi, T., & Buckley, B. (2012). Customer Churn Prediction in Telecommunications. Expert Systems with Applications.
[14] Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). A Profit-Driven Data Mining Approach for Churn Prediction. European Journal of Operational Research.
[15] Keramati, A., Ghaneei, H., & Mirmohammadi, S. M. (2016). Developing a Prediction Model for Customer Churn Using Data Mining. Financial Innovation.