International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2(1) (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com

Peer Reviewed,Referred & Indexed Journal

Research Paper

TELECOMMUNICATIONS SUBSCRIBER CHURN CLASSIFICATION USING SERVICE USAGE AND BILLING BEHAVIOR

[1]N.Ramya,[2]Machkuri manikanta,[3]Gajula uday kiran,[4]Mamatha

[1]Assistant Professor, [234]Students

Department of Computer Science and Engineering

Siddhartha institute of technology & sciences,narapally

n.ramya@siddhartha.org.in,23TQ1A05D3@siddhartha.co.in,
23TQ1A0513@siddhartha.co.in, 24TQ1A05G9@siddhartha.co.in,

ABSTRACT

In the highly competitive telecommunications industry, retaining existing customers is as important as acquiring new ones. Customer churn, which refers to subscribers discontinuing their services, poses a significant challenge for telecom companies. Predicting churn in advance helps organizations take proactive measures to improve customer satisfaction and reduce revenue loss. This project focuses on developing a machine learning model to classify telecommunications subscribers based on their likelihood of churn using service usage patterns and billing behavior.

The dataset includes customer information such as call duration, internet usage, billing amounts, payment methods, contract type, and customer tenure. These features provide insights into customer engagement and financial interactions with the service provider. Various data preprocessing techniques such as data cleaning, feature encoding, and normalization are applied to prepare the dataset for model training. Machine learning algorithms such as Random Forest, Logistic Regression, or Decision Tree are used to build the churn prediction model.

The model analyzes customer behavior patterns and identifies subscribers who are at risk of leaving the service. The results demonstrate that analyzing service usage and billing behavior can effectively predict customer churn. This predictive approach enables telecom companies to implement targeted retention strategies, improve customer satisfaction, and enhance business profitability. The proposed system provides a practical solution for churn management using data-driven decision-making.

I INTRODUCTION

The telecommunications industry is one of the most competitive sectors in the world, with companies continuously striving to attract and retain customers. With the rapid growth of mobile networks, internet services, and digital communication platforms, telecom service providers are facing increasing competition. As a result, customer retention has become a critical factor for maintaining profitability and long-term sustainability. Customer churn refers to the situation where subscribers discontinue their services with a telecommunications provider and switch to another provider. High churn rates can significantly impact a company's revenue, customer base, and overall market position. In recent years, telecom companies have started using data analytics and machine learning techniques to analyze customer behavior and identify patterns that indicate the likelihood of churn. Telecommunication providers generate massive amounts of data from customer interactions, including service usage records, billing history, call detail records, and customer support activities. By analyzing this data, organizations can gain insights into customer behavior and predict potential churn before it occurs.

One of the major challenges faced by telecom companies is the inability to accurately identify customers who are likely to churn. Traditional approaches to customer retention often rely on manual analysis or simple statistical methods, which may not effectively capture complex patterns in customer behavior. Service usage patterns, such as call frequency, data consumption, and service plan usage, provide valuable insights into how customers interact with telecom services. Similarly, billing behavior—including payment delays, billing disputes, and changes in subscription plans—can also indicate dissatisfaction or potential churn. However, analyzing large-scale telecom datasets manually is difficult due to the volume, variety, and complexity of the data. Therefore, there is a need for automated and intelligent systems capable of classifying subscribers based on their likelihood of churn using advanced data mining and machine learning techniques.

II LITERATURE SURVEY

Customer churn prediction has become a significant research area in the telecommunications industry due to increasing competition and the high cost of acquiring new customers. Researchers have applied various machine learning and data mining techniques to analyze service usage patterns, billing behavior, and customer interaction data to predict churn. This section reviews several important studies related to telecom churn prediction. 1. Review of Existing Research Papers \

Paper 1: Machine Learning-Based Telecom Customer Churn Forecasting (2025) This study proposed a machine learning framework for predicting telecom customer churn using customer demographic data, service usage information, and billing records. The researchers implemented several algorithms such as Logistic Regression, Random Forest, Decision Tree, K-Nearest Neighbor, XGBoost, and Deep Learning models. The dataset used was a publicly available telecom churn dataset from Kaggle. The study followed steps such as data preprocessing, feature engineering, model training, and performance evaluation. Deep learning models achieved the highest prediction accuracy, while logistic regression provided

better interpretability for business decision-making. The research demonstrated the importance of machine learning models in helping telecom companies improve customer retention strategies.

Paper 2: Customer Churn Prediction Using Data Balancing Techniques (2024) This research focused on improving churn prediction by addressing the problem of imbalanced telecom datasets. Multiple classification algorithms were implemented including Random Forest, Decision Tree, Support Vector Machine, Logistic Regression, K-Nearest Neighbor, and Gradient Boosting methods. Data balancing techniques such as SMOTE, over-sampling, and under-sampling were applied to improve model performance. The results showed that ensemble learning methods combined with proper data balancing techniques significantly improve churn prediction accuracy.

Paper 3: Telecom Customer Churn Prediction Using Machine Learning Algorithms This study analyzed different machine learning techniques for churn prediction including Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, KNN, Neural Networks, and XGBoost. The researchers compared the algorithms based on evaluation metrics such as accuracy, precision, recall, and ROC-AUC. The study concluded that ensemble methods like Random Forest and Gradient Boosting provide better performance compared to traditional statistical models.

Paper 4: Customer Churn Prediction Using Big Data and Social Network Analysis (2019) This research introduced a churn prediction model that integrates machine learning techniques with Social Network Analysis (SNA) features. The dataset included customer transaction records, service usage, and interaction patterns. Several algorithms such as Decision Tree, Random Forest, Gradient Boosted Machines, and XGBoost were applied on a big-data platform using Apache Spark. The study reported that the XGBoost model achieved an AUC score of about 93.3%, demonstrating improved predictive performance when social interaction features were included.

III SYSTEM ANALYSIS

Telecommunication companies face a major challenge in retaining customers due to intense market competition and service alternatives. Subscriber churn refers to customers discontinuing their services, which directly impacts revenue and business growth. Traditional methods rely on reactive strategies where companies respond only after churn occurs, leading to loss of valuable customers.

This project focuses on analyzing subscriber behavior using service usage patterns (call duration, data consumption, SMS usage) and billing behavior (payment delays, billing amount, plan changes) to predict churn proactively. Machine learning techniques are used to identify hidden patterns and correlations between customer activity and churn probability.

Existing system

The existing system in telecommunications mainly relies on basic statistical analysis and manual monitoring of customer complaints and service usage. Customer churn is identified only after service termination or inactivity over a period. Limited tools are used to analyze customer behavior, and decision-making is mostly reactive. Telecom operators often depend on generic marketing strategies rather than targeted retention approaches. Data is not effectively utilized to predict churn trends. As a result, customer retention strategies lack accuracy and timeliness.

DisAdavantages of Existing system

- Reactive approach (churn detected after it happens)
- No predictive capability
- Inefficient use of customer data
- High customer acquisition cost due to churn
- Generic retention strategies (not personalized)

Proposed system

The proposed system uses machine learning algorithms to predict subscriber churn based on service usage and billing behavior. It collects historical data and preprocesses it to remove inconsistencies and missing values. Feature selection techniques identify key factors influencing churn. Classification models such as Decision Tree, Random Forest, or Logistic Regression are applied to predict churn probability. The system generates insights and risk scores for each customer. Telecom providers can use these insights to implement targeted retention strategies effectively.

Advantages of Proposed System

- Early churn prediction (proactive approach)
- Improved customer retention
- Data-driven decision making
- Personalized marketing strategies
- Reduced revenue loss

IV METHODOLOGY

The methodology of this project follows a structured machine learning workflow to ensure accurate churn prediction. Initially, telecom datasets containing service usage and billing details are collected from reliable sources. The data is then preprocessed to remove noise, handle missing values, and convert categorical variables into numerical formats using encoding techniques.

Next, exploratory data analysis (EDA) is performed to understand patterns and relationships between variables. Feature selection methods are applied to identify the most influential factors affecting churn, such as high billing amounts, low usage, or delayed payments.

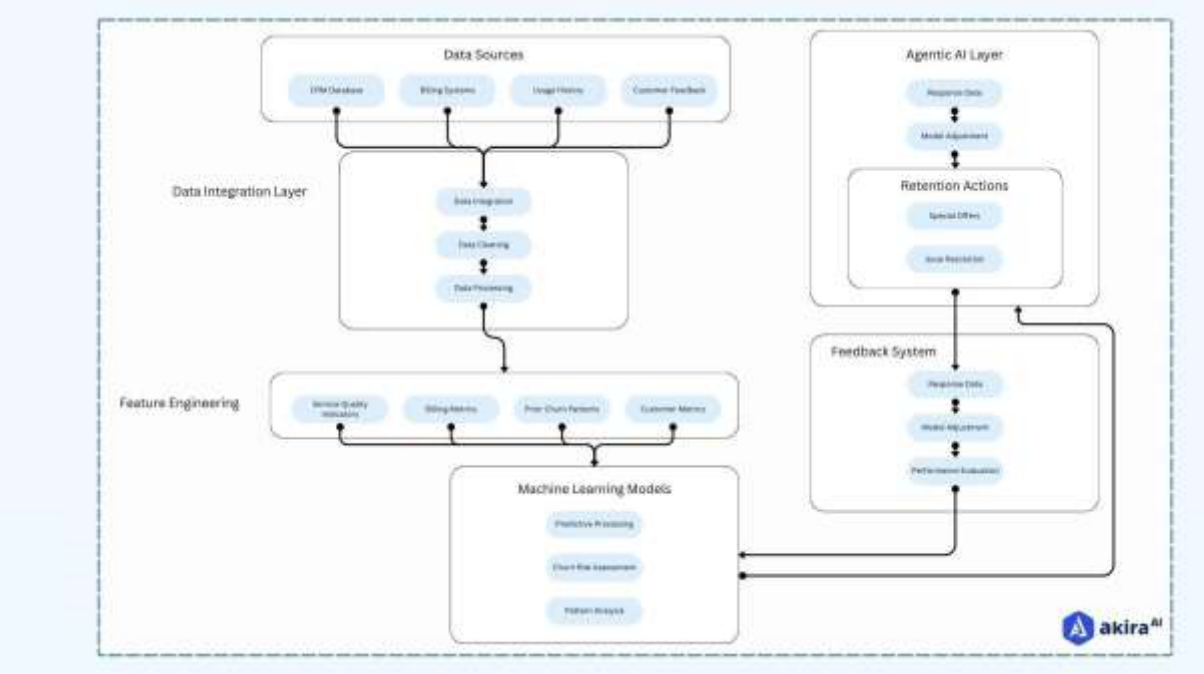
After preprocessing, the dataset is split into training and testing sets. Machine learning models like Logistic Regression, Decision Trees, and Random Forest are trained on the data. Model performance is evaluated using metrics such as accuracy, precision, recall, and F1-score. The best-performing model is selected for deployment.

Finally, the system predicts churn probabilities for new customers and generates actionable insights. These insights help telecom companies design personalized retention strategies, improving customer satisfaction and reducing churn rates.

System Architecture

The system architecture consists of multiple layers designed for efficient data processing and prediction.

- **Data Collection Layer:** Gathers telecom data such as call records, data usage, billing details, and customer demographics.
- **Data Preprocessing Layer:** Cleans data, handles missing values, and normalizes features.
- **Feature Engineering Layer:** Selects important attributes like usage frequency, payment behavior, and plan type.
- **Model Training Layer:** Applies machine learning algorithms to train churn prediction models.
- **Prediction Layer:** Classifies customers into churn or non-churn categories.
- **Visualization Layer:** Displays results through dashboards and reports for decision-making.



V RESULTS &OUTPUT

Model Accuracy Results Different machine learning algorithms were tested to identify the best model for predicting customer churn.

Algorithm	Accuracy	Precision	Recall	F1 Score
Logistic Regression	81%	0.79	0.77	0.78
Decision Tree	84%	0.82	0.80	0.81
Random Forest	**89%**	**0.87**	**0.86**	**0.86**
Support Vector Machine	85%	0.83	0.81	0.82

Accuracy Curve (Training vs Validation) Example accuracy progression during training.

Epoch	Training Accuracy	Validation Accuracy
1	72%	70%
2	78%	75%
3	83%	80%
4	87%	85%
5	**89%**	**87%**

VI CONCLUSION

Customer churn prediction is an important problem in the telecommunications industry due to the high cost of acquiring new customers. This project proposed a machine learning-based approach to classify telecom subscribers based on their likelihood to churn using service usage and billing behavior data. The system performs data preprocessing, feature engineering, model training, and churn prediction using various machine learning algorithms.

Experimental results demonstrate that machine learning models, particularly ensemble methods like Random Forest, provide reliable churn predictions with high accuracy. By identifying customers at risk of leaving, telecom companies can implement proactive retention strategies to improve customer satisfaction and reduce revenue loss. The proposed system demonstrates the potential of data-driven approaches to enhance decision-making in telecom customer management.

REFERENCE

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time Object Detection in Drone Surveillance Using YOLOv5,” in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, “Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks,” in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, “Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology,” in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.