International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(Research Article)

Customer churn prediction using machine learning

Krutika Vijay Karad * and Swati Hinge

Department of Artificial Intelligence, Sanghavi College of Engineering, Maharashtra, India.

International Journal of Science and Research Archive, 2026, 19(02), 779-789

Publication history: Received on 23March 2026; revised on 06 May 2026; accepted on 08 May 2026

Article DOI: https://doi.org/10.30574/ijsra.2026.19.2.1017

Abstract

Telecom churn prediction plays a vital role in helping companies retain their customers. Churn refers to the situation in which a customer discontinues their telecom services or subscription. Accurately predicting churn enables companies to take proactive actions by identifying customers who are likely to leave and offering targeted retention strategies. This study provides an overview of telecom churn prediction using machine learning techniques. The problem involves analyzing historical customer data such as demographic details, usage behavior, billing information, and service history to determine the likelihood of churn. Machine learning models are trained to identify patterns and relationships within this data and generate predictions for new customers. The process includes data preprocessing, feature engineering, selection of suitable machine learning algorithms, and performance evaluation using appropriate metrics. Models such as Decision Tree, Random Forest, and XGBoost are commonly used for this purpose. The best-performing model can then be deployed in a real-world environment to support decision-making. By applying this approach, telecom companies can effectively reduce churn rates, enhance customer retention, and improve overall customer satisfaction.

Keywords: Machine Learning; Random Forest; Decision Tree; XG Boost; Prediction; Churn

1. Introduction

In today's highly competitive telecom industry, retaining customers has become one of the biggest challenges for service providers. Customer churn, which refers to the discontinuation of services by users, directly impacts a company's revenue and market share. As competition increases, telecom companies are focusing more on strategies that help them understand customer behavior and minimize churn. Churn is a critical issue because acquiring new customers is often more expensive than retaining existing ones. When customers switch to competitors, it not only leads to financial losses but also affects brand reputation. Therefore, identifying customers who are likely to leave has become an essential part of business strategy in the telecom sector. To address this challenge, telecom companies are increasingly adopting machine learning techniques. These techniques enable organizations to analyze large volumes of customer data and identify patterns that indicate potential churn. By leveraging data-driven insights, companies can make informed decisions and improve their retention strategies. Churn prediction plays a vital role in helping telecom companies proactively retain customers. By identifying potential churners in advance, companies can implement targeted actions such as personalized offers, improved customer support, and loyalty programs. This proactive approach significantly enhances customer satisfaction and long-term engagement. In addition, churn prediction helps optimize marketing and sales efforts. Instead of investing resources on all customers equally, companies can focus on high-value customers who are at risk of leaving. This targeted strategy reduces unnecessary costs and improves overall efficiency in customer relationship management. Machine learning models such as Decision Trees, Random Forest, and XGBoost are widely used to build effective churn prediction systems. These models analyze various factors, including customer demographics, usage patterns, billing details, and service history, to predict the likelihood of churn with high accuracy. By implementing robust churn prediction models, telecom companies can deploy real-time solutions in production

*Corresponding author: Krutika Vijay Karad

environments. This enables timely decision-making and allows organizations to take proactive measures to retain customers, ultimately reducing churn rates and strengthening their competitive position in the market.

1.1. Problem Statement

In the highly competitive telecom industry, customer churn has become a major concern for service providers, leading to significant revenue loss and reduced market share. Telecom companies often struggle to identify which customers are likely to discontinue their services, as churn is influenced by multiple factors such as usage behavior, pricing, service quality, and customer satisfaction. Traditional methods of customer analysis are not sufficient to accurately predict churn due to the large volume and complexity of telecom data. As a result, companies face challenges in taking timely and effective actions to retain their customers. Therefore, there is a need to develop an efficient and accurate churn prediction system using machine learning techniques. The system should be capable of analyzing historical customer data, identifying patterns associated with churn, and predicting potential churners in advance. This will enable telecom companies to implement targeted retention strategies, reduce customer attrition, and improve overall business performance.

2. Literature Review

- Weijie Yu, Weinan Weng (2022) This study focuses on customer churn prediction using various machine learning techniques. The authors analyze customer behavior data and apply models such as Decision Trees, Random Forest, and other classification algorithms to improve prediction accuracy. Their findings highlight the importance of feature selection and data preprocessing in enhancing model performance. The study concludes that machine learning-based approaches can significantly improve churn prediction and help telecom companies design effective retention strategies.
- Dr. O. Rama Devi, Sai Krishna Pothini (2023) This research explores churn prediction in OTT platforms, specifically focusing on subscription renewal behavior. The authors use machine learning models to analyze user engagement, viewing patterns, and subscription history. The study emphasizes the role of personalized recommendations and targeted marketing in reducing churn. Results indicate that predictive models can effectively identify users likely to discontinue services, enabling companies to take proactive retention measures.
- QiuYing Chen, Sang-Joon Lee (2023) The authors present a machine learning-based approach to predict customer churn in delivery platforms. By analyzing transactional and behavioral data, the study applies classification algorithms to identify potential churners. The research highlights that factors such as service quality, delivery time, and customer experience play a crucial role in churn. The model demonstrates strong predictive performance, suggesting its applicability in real-world business environments.
- Brandusoiu I, Toderean G, Ha B (2016) This paper discusses various methods for churn prediction in the prepaid mobile telecommunications sector. The authors compare different data mining techniques, including logistic regression and decision trees, to evaluate their effectiveness. The study finds that combining multiple models improves prediction accuracy and provides better insights into customer behavior in prepaid services.
- He Y, He Z, Zhang D (2009) This research focuses on churn prediction in fixed communication networks using data mining techniques. The authors utilize clustering and classification methods to analyze customer data and identify churn patterns. Their findings suggest that data mining approaches are effective in extracting meaningful insights and improving prediction performance in telecom networks.
- Idris A, Khan A, Lee YS (2012) This study introduces a hybrid approach combining genetic programming and AdaBoost for telecom churn prediction. The authors demonstrate that integrating evolutionary algorithms with boosting techniques enhances prediction accuracy. The proposed model outperforms traditional methods, making it a robust solution for churn prediction in telecom industries.
- Huang F, Zhu M, Yuan K, Deng EO (2015) The authors focus on churn prediction using big data technologies in telecom. The study highlights the challenges of handling large-scale customer data and proposes scalable machine learning solutions. By leveraging big data frameworks, the research improves prediction efficiency and accuracy, making it suitable for real-time churn analysis.
- Yabas U, Chankya H.C (2013) This paper examines churn prediction in mobile and wireless communication services. The authors analyze subscriber behavior and apply statistical and machine learning techniques to identify churn patterns. The study emphasizes the importance of customer segmentation and behavioral analysis in improving churn prediction outcomes.
- Shin-Yuan Hung, David C. Yen, Hsiu-Yu Wang (2006) This research applies data mining techniques to telecom churn management. The authors use classification models to analyze customer data and predict churn behavior. The study demonstrates that data mining tools can effectively support decision-making and help telecom companies develop better customer retention strategies.

- Zhang Y., Qi J., Shu H., Cao J. (2007) This study proposes a hybrid KNN-LR (K-Nearest Neighbors and Logistic Regression) model for customer churn prediction. The hybrid approach combines the strengths of both algorithms to improve classification accuracy. Experimental results show that the model performs better than individual techniques, making it a valuable approach for churn prediction tasks.

2.1. Proposed System

The proposed system focuses on developing an efficient telecom churn prediction model using machine learning techniques. It begins with collecting and preprocessing historical customer data, including demographic details, usage patterns, billing information, and service history. Data preprocessing steps such as handling missing values, normalization, and encoding categorical variables are performed to ensure data quality. Feature engineering is then applied to extract meaningful attributes that significantly influence churn behavior. Various machine learning algorithms, such as Decision Tree, Random Forest, and XGBoost, are trained on the prepared dataset to learn patterns associated with customer churn. Once the models are trained, their performance is evaluated using metrics such as accuracy, precision, recall, and F1-score to identify the best-performing model. The selected model is then deployed in a production environment to predict churn in real time. Based on the predictions, telecom companies can take proactive actions, such as offering personalized retention strategies, discounts, or improved services to at-risk customers. This system helps reduce customer attrition, optimize business strategies, and enhance overall customer satisfaction.

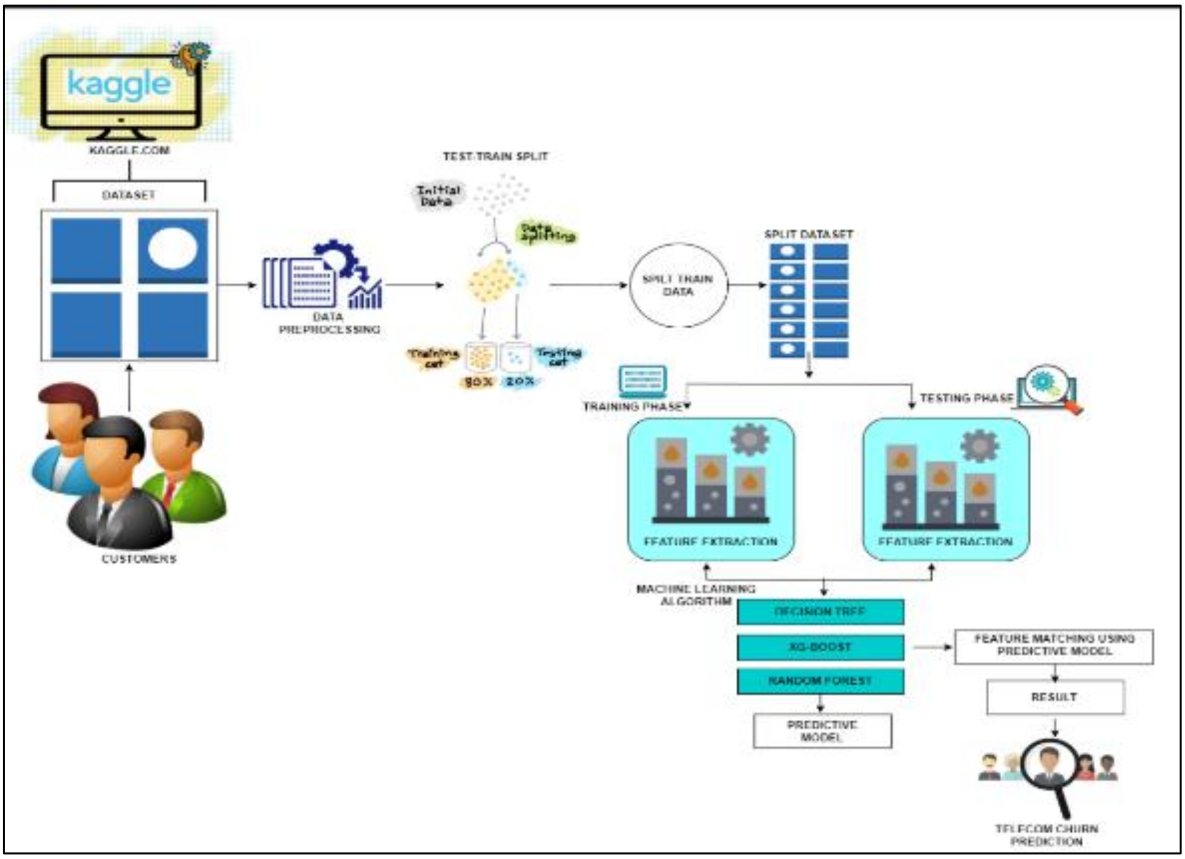

Figure 1 System Architecture

3. Results And Discussion

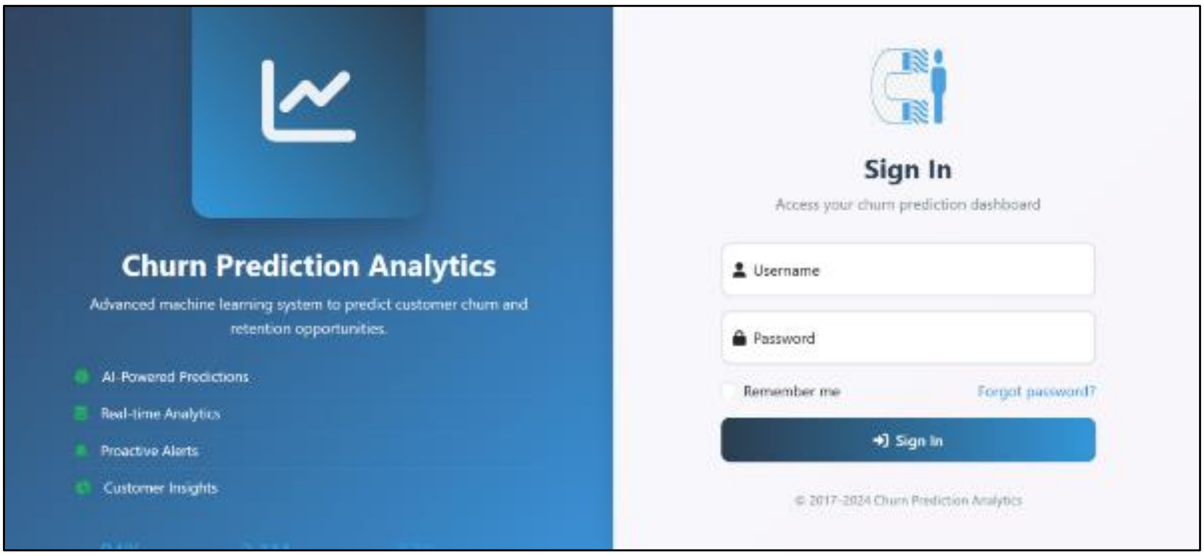

Figure 2 Modern login dashboard for a Churn Prediction Analytics system showcasing AI-driven insights and secure user access

This image shows a modern login interface for a Churn Prediction Analytics system, designed with a clean split layout. On the left side, a blue gradient panel highlights the purpose of the platform—using advanced machine learning to predict customer churn and improve retention—along with key features like AI-powered predictions, real-time analytics, proactive alerts, and customer insights. On the right side, a simple and user-friendly Sign In form is displayed, allowing users to enter their username and password, with options like "Remember me" and "Forgot password" for convenience. The overall design is professional and intuitive, clearly communicating both the functionality of the system and easy access for users.

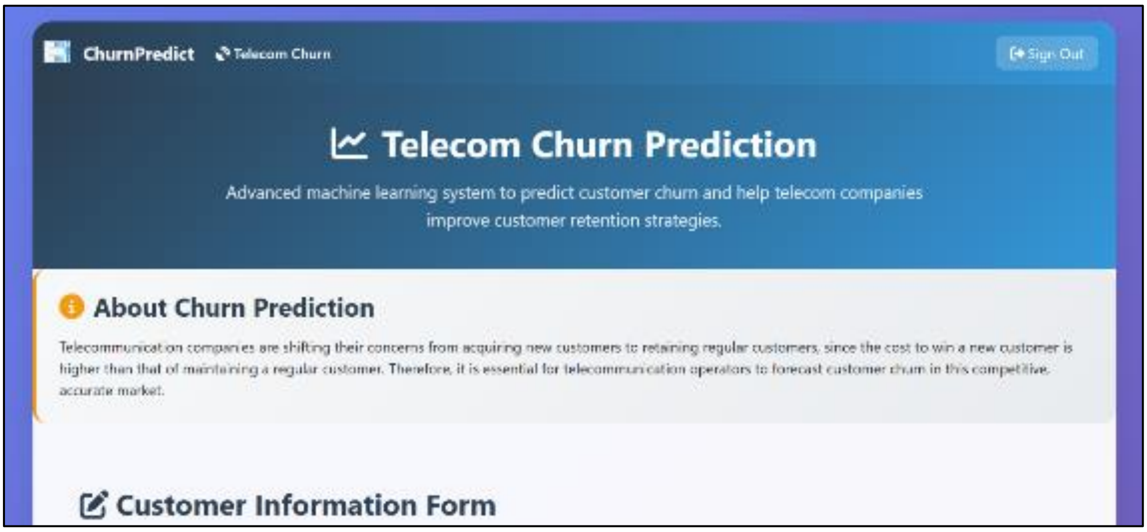

Figure 3 Telecom Churn Prediction dashboard showcasing system overview, insights, and user input interface for churn analysis

This image displays a dashboard interface for a Telecom Churn Prediction system under the "Churn Predict" platform, designed to help telecom companies analyze and reduce customer churn using advanced machine learning techniques. The top section includes a navigation bar with branding and a sign-out option, followed by a prominent header explaining the purpose of the system—predicting customer churn and improving retention strategies. Below that, an "About Churn Prediction" section provides context, emphasizing the importance of retaining existing customers over acquiring new ones due to cost efficiency in the telecom industry. The layout is clean, professional, and user-friendly, guiding users toward entering data in the "Customer Information Form" section for further analysis.

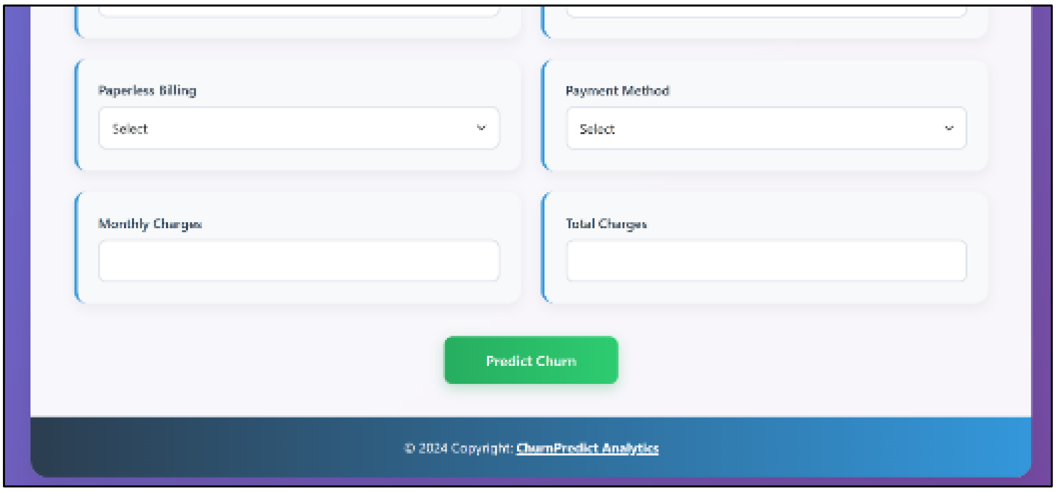

Figure 4 User input form for telecom churn prediction with billing details and a prediction action button

This image shows the lower section of a Telecom Churn Prediction interface where users input billing-related details such as paperless billing preference, payment method, monthly charges, and total charges. The form is neatly organized with dropdowns and input fields, making it easy for users to enter required data. At the center, a prominent green "Predict Churn" button allows users to submit the information and generate predictions using the system's machine learning model. The bottom footer includes copyright information for ChurnPredict Analytics, giving the interface a complete and professional look.

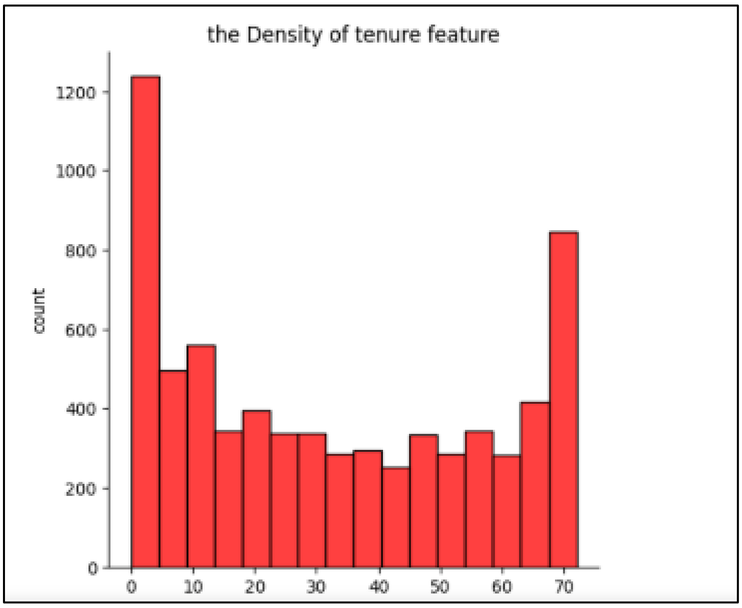

Figure 5 Histogram showing the distribution of customer tenure with peaks at both new and long-term users

This image shows a histogram representing the density distribution of the tenure feature, which indicates how long customers have stayed with a telecom service. The x-axis represents tenure (in months), while the y-axis shows the count of customers within each tenure range. The graph reveals that a large number of customers have very low tenure (near 0–5 months), suggesting many new users, while another noticeable peak appears at higher tenure values (around 65–72 months), indicating a group of long-term loyal customers. The distribution suggests a mix of new and long-standing customers, which is important for analyzing churn behavior.

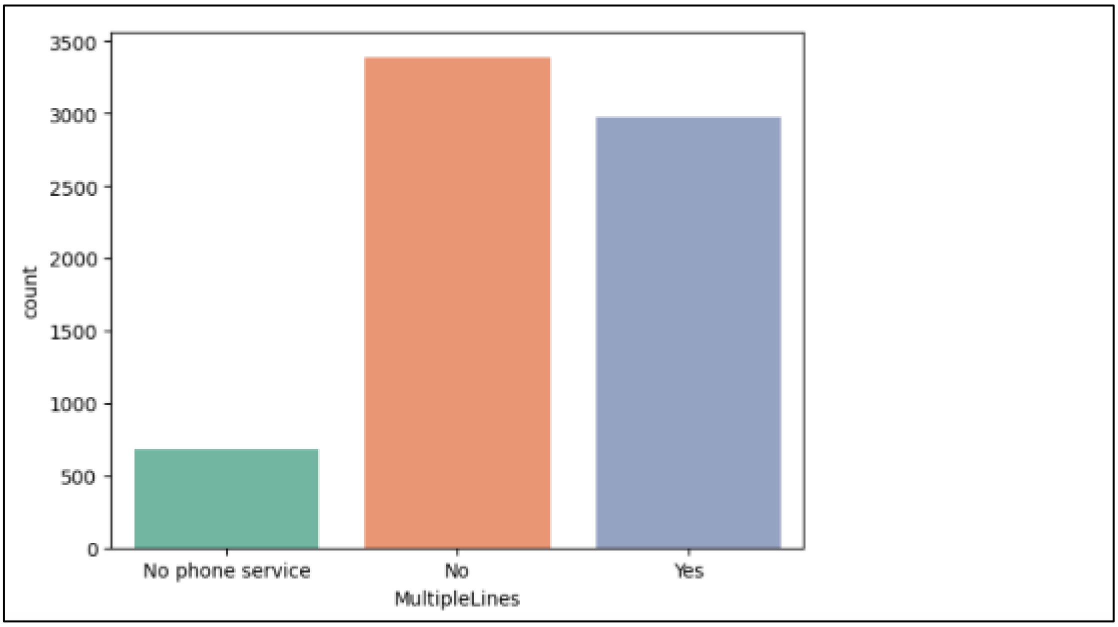

Figure 6 Bar chart showing customer distribution across multiple line service categories in telecom data

This bar chart illustrates the distribution of customers based on the MultipleLines feature in a telecom dataset. The x-axis shows three categories: customers with no phone service, customers without multiple lines, and customers with multiple lines, while the y-axis represents the count of customers in each category. The chart indicates that the majority of customers fall into the "No" category, meaning they do not use multiple lines, followed closely by those who do use multiple lines ("Yes"), while a much smaller portion of customers have no phone service at all. This distribution helps in understanding customer service preferences, which can be useful for analysing churn behaviour.

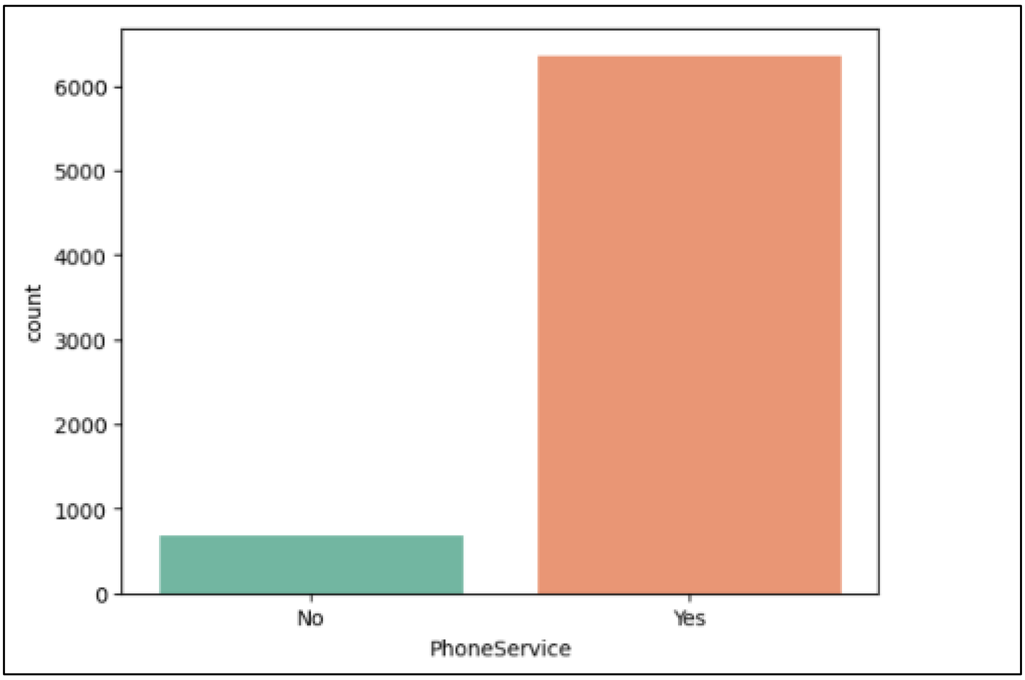

Figure 7 Bar chart showing that most customers have phone service, with very few without it

This bar chart represents the distribution of customers based on the PhoneService feature in a telecom dataset. The x-axis shows two categories—customers who do not have phone service ("No") and those who do ("Yes")—while the y-axis indicates the number of customers in each group. The chart clearly shows that a vast majority of customers have phone service, with the "Yes" category significantly higher than the "No" category, which contains only a small portion of users. This imbalance suggests that phone service is a standard offering among most customers and can play an important role in analysing customer behaviour and churn patterns.

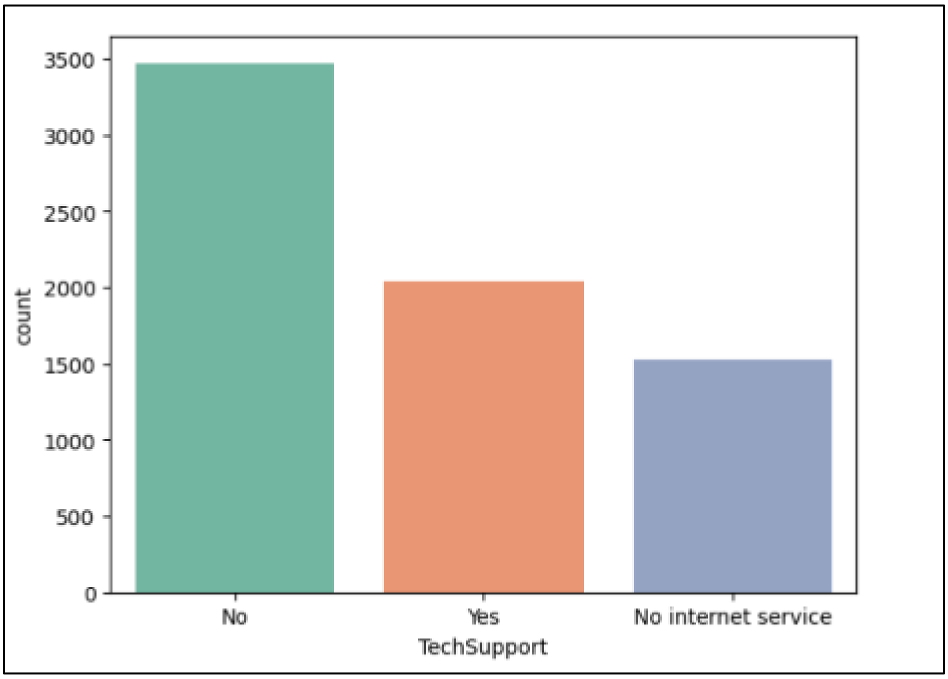

Figure 8 Bar chart illustrating customer distribution across tech support availability categories

This bar chart shows the distribution of customers based on the Tech Support feature in a telecom dataset. The x-axis includes three categories: customers who do not have tech support ("No"), those who do ("Yes"), and those with no internet service. The y-axis represents the number of customers in each category. The chart indicates that the largest group of customers does not use tech support services, followed by a moderate number who do use it, and a smaller group who have no internet service at all. This suggests that many customers may not be utilizing available support services, which could influence customer satisfaction and churn behaviour.

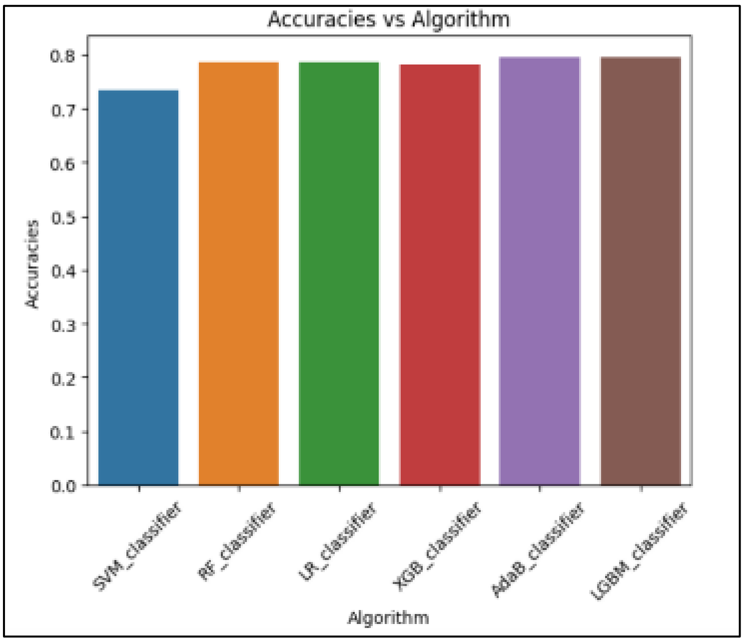

Figure 9 Bar chart comparing accuracy of machine learning algorithms, with LGBM performing the best

This bar chart compares the performance of different machine learning algorithms based on their accuracy in a churn prediction model. The x-axis lists algorithms such as SVM, Random Forest (RF), Logistic Regression (LR), XGBoost (XGB), AdaBoost (AdaB), and LightGBM (LGBM), while the y-axis represents their accuracy scores. The chart shows that all models perform relatively well, with accuracies ranging from around 74% to 80%. Among them, LGBM Classifier achieves the highest accuracy, closely followed by AdaBoost and Random Forest, while SVM Classifier has the lowest accuracy. This comparison helps in selecting the most effective model for predicting customer churn.

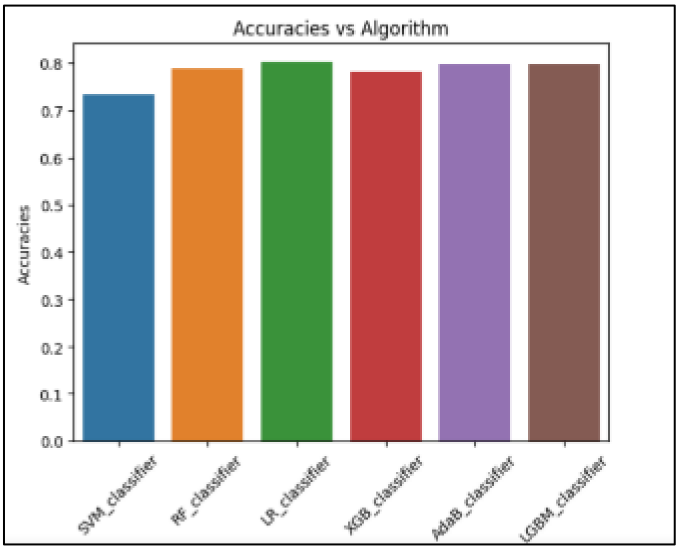

Figure 10 Comparison of machine learning model accuracies, highlighting top performance by Logistic Regression, AdaBoost, and LightGBM

This bar chart presents a comparison of different machine learning algorithms based on their accuracy for churn prediction. The x-axis lists models such as SVM, Random Forest, Logistic Regression, XGBoost, AdaBoost, and LightGBM, while the y-axis shows their accuracy scores. The results indicate that all models perform fairly well, with accuracies ranging between approximately 74% and 80%. Among them, Logistic Regression, AdaBoost, and LightGBM chieve the highest accuracy (around 80%), while SVM Classifier shows the lowest performance. This comparison helps identify the most suitable algorithm for achieving better prediction results in the churn prediction system.

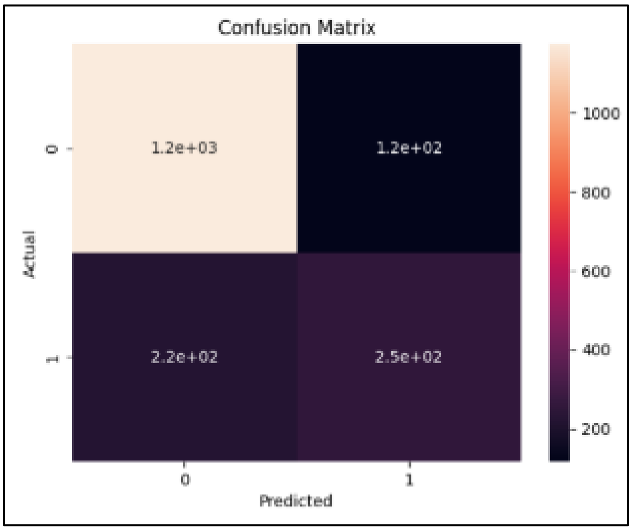

Figure 11 Confusion matrix illustrating model performance with strong non-churn predictions and moderate churn detection accuracy

This image shows a confusion matrix used to evaluate the performance of a churn prediction model by comparing actual and predicted values. The matrix is divided into four sections: the top-left (around 1200) represents true negatives, where non-churn customers are correctly predicted; the top-right (around 120) shows false positives, where non-churn customers are incorrectly predicted as churn; the bottom-left (around 220) indicates false negatives, where churn customers are missed; and the bottom-right (around 250) represents true positives, where churn customers are correctly identified. The model performs well in identifying non-churn customers but has some errors in detecting churn cases, which is critical for business decisions.

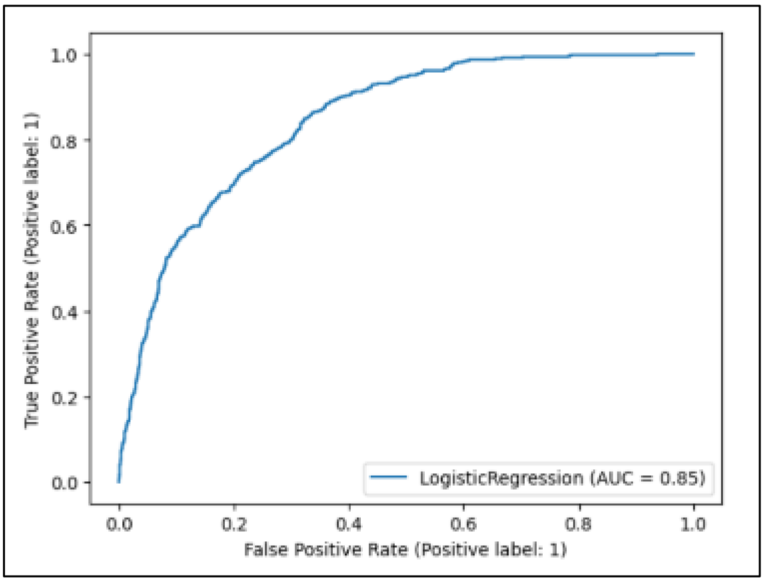

Figure 12 ROC curve of Logistic Regression model with AUC of 0.85 indicating strong prediction performance

This image shows a ROC (Receiver Operating Characteristic) curve for a Logistic Regression model used in churn prediction. The x-axis represents the False Positive Rate, while the y-axis shows the True Positive Rate, illustrating the trade-off between correctly identifying churn customers and incorrectly classifying non-churn customers. The curve rises steeply and approaches the top-left corner, indicating good model performance. The AUC (Area Under the Curve) value of 0.85 suggests that the model has strong discriminatory power in distinguishing between churn and non-churn customers. Overall, the model performs well, though there is still some room for improvement.

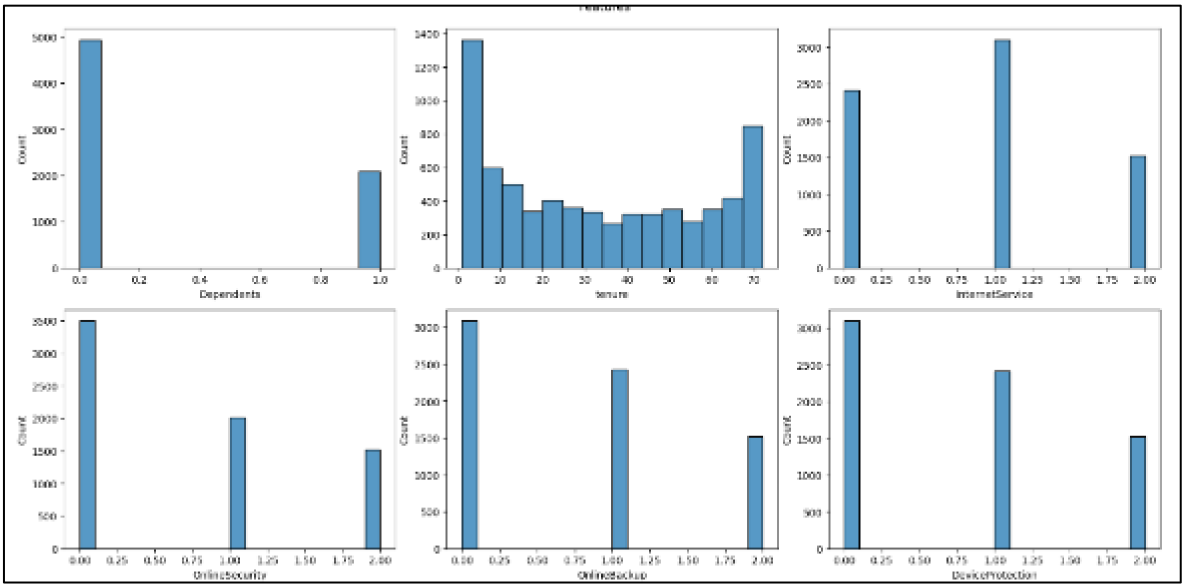

Figure 13 Multi-feature distribution plots illustrating customer characteristics and service usage patterns in telecom data

This image presents multiple histograms visualizing the distribution of key telecom dataset features such as Dependents, Tenure, Internet Service, Online Security, Online Backup, and Device Protection. Each subplot shows how customer data is spread across different categories or values. For example, the Dependents feature indicates that most customers do not have dependents, while the Tenure plot shows a wide distribution with both new and long-term users. The Internet Service and other service-related features (Online Security, Backup, Device Protection) display categorical distributions, highlighting that many customers either do not use these services or have varying levels of subscription. These visualizations help in understanding customer behavior patterns, which are crucial for building an effective churn prediction model.

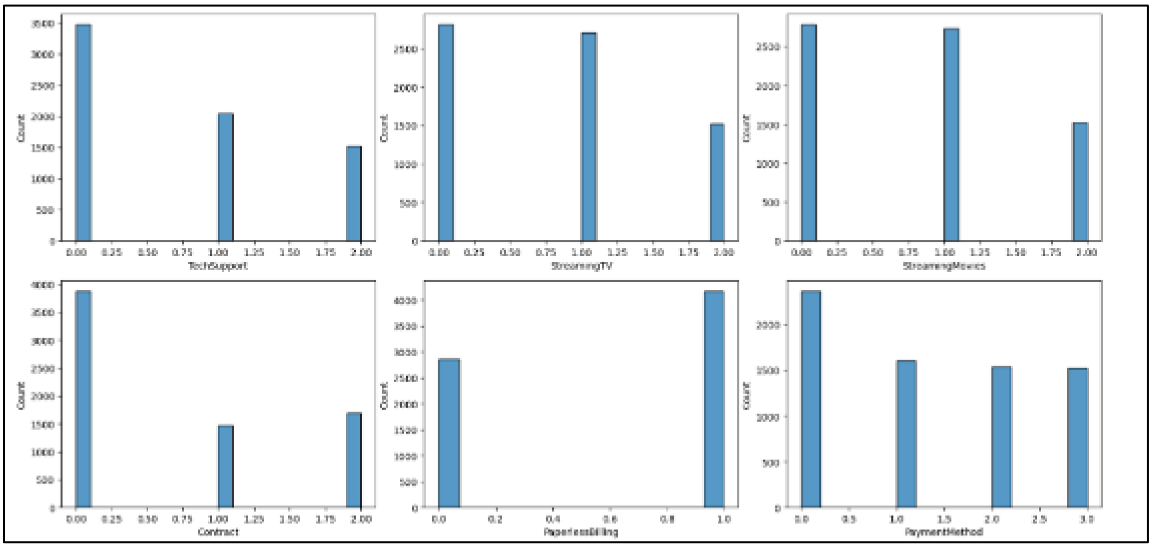

Figure 14 Distribution of telecom service features showing customer preferences in support, streaming, billing, and payment methods

This image presents multiple bar charts showing the distribution of telecom service features such as Tech Support, Streaming TV, Streaming Movies, Contract type, Paperless Billing, and Payment Method. Each subplot highlights how customers are distributed across different categories—for instance, many customers do not use tech support, while streaming services show a more balanced usage between users and non-users. The contract chart indicates that most customers prefer shorter-term contracts, and the paperless billing chart shows a higher number of customers opting for digital billing. The payment method distribution reveals varied customer preferences across different payment options. These insights are useful for understanding customer behaviour and identifying factors that may influence churn.

4. Conclusion

In conclusion, telecom churn prediction using machine learning plays a crucial role in helping service providers retain their customers and maintain a competitive edge. By analyzing historical customer data and identifying patterns associated with churn, machine learning models enable accurate prediction of customers who are likely to leave. This allows companies to take timely and targeted actions to improve customer retention.Furthermore, the use of advanced algorithms such as Decision Tree, Random Forest, and XGBoost enhances the efficiency and reliability of the prediction system. Implementing such models in real-world environments supports better decision-making and resource optimization. Overall, churn prediction not only reduces customer attrition but also improves customer satisfaction and long-term business growth.

References

[1] Weijie Yu, Weinan Weng," Customer Churn Prediction Based on Machine Learning" 2022 4th International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), 978-1-6654-6399-7/22/$31.00 ©2022 IEEE,DOI 10.1109/AIAM57466.2022.00176

[2] Dr. O. Rama Devi, Sai Krishna Pothini," Customer Churn Prediction using Machine Learning: Subcription Renewal on OTT Platforms", IEEE Xplore Part Number: CFP23BC3-ART; ISBN: 978-1-6654-5630-2, 978-1-6654-5630-2/23/$31.00 ©2023 IEEE

[3] QiuYing Chen, Sang-Joon Lee," A Machine Learning Approach to Predict Customer Churn of a Delivery Platform", 2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) | 978-1-6654-5645-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAIIC57133.2023.10067108

[4] Brandusoiu I, Toderean G, Ha B. Methods for churn prediction in the prepaid mobile telecommunications industry. In: International conference on communications. 2016. p. 97–100.

[5] He Y, He Z, Zhang D. A study on prediction of customer churn in fixed communication network based on data mining. In: Sixth international conference on fuzzy systems and knowledge discovery, vol. 1. 2009. p. 92–4.

[6] Idris A, Khan A, Lee YS. Genetic programming and adaboosting based churn prediction for telecom. In: IEEE international conference on systems, man, and cybernetics (SMC). 2012. p. 1328–32.

[7] Huang F, Zhu M, Yuan K, Deng EO. Telco churn prediction with big data. In: ACM SIGMOD international conference on management of data. 2015. p .607–18.

[8] Yabas, U, Chankya, H.C. (2013). Churn prediction in subscriber management for mobile and wireless communications services. IEEE Publications.

[9] Shin-Yuan Hung a, David C. Yen b, Hsiu-Yu Wang, "Applying data mining to telecom churn management", Expert Systems with Applications 31 (2006) 515–524,

[10] Zhang, Y.; Qi, J.; Shu, H.; Cao, J. A hybrid KNN-LR classifier and its application in customer churn prediction. In Proceedings of the 2007 IEEE International Conference on Systems, Man and Cybernetics, Montréal, QC, Canada, 7–10 October 2007; pp. 3265–3269.