International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 02 | Feb - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Customer Churn Prediction Using Machine Learning

Amit Talele[1], Sarthak Pawale[2], Adarsh Sonawane[3], Yogesh Salunkhe[4]

[1]Amit Talele Department of Information Technology From Matoshri Aasarabai Polytechnic.
[2]Sarthak Pawale Department of Information Technology From Matoshri Aasarabai Polytechnic.
[3]Adarsh Sonawane Department of Information Technology From Matoshri Aasarabai Polytechnic.
[4]Yogesh Salunkhe Department of Information Technology From Matoshri Aasarabai Polytechnic.
[5]Mrs. Pratiksha Gadakh Lecturer of Information Technology From Matoshri Aasarabai Polytechnic.
[6]Mr. Mahesh Bhandakkar Head of Information Technology From Matoshri Aasarabai Polytechnic.

Abstract - Customer attrition poses a major issue for the telecommunications sector, resulting in revenue decline and higher expenses for gaining new customers. Predicting churn accurately enables telecom companies to take proactive measures to retain customers. This study leverages specifically Random Forest, Decision Tree, and XGBoost, to develop a robust churn prediction model. These algorithms analyze customer behavioral data, including call duration, internet usage, billing history, and complaints, to identify potential churners. Decision Tree provides an interpretable model, Random Forest improves predictive accuracy through ensemble learning, and XGBoost enhances performance with gradient boosting and optimized handling of imbalanced datasets. The proposed model assists telecom companies in classifying customers based on their churn risk, enabling the implementation of targeted retention approaches like tailored discounts and rewards programs. By integrating advanced machine learning techniques, telecom service providers can enhance customer retention, minimize the churn rates, and improves the business sustainability. The study highlights the importance of data-driven decision-making in the telecom sector, demonstrating how predictive analytics can optimize customer relationship management and drive profitability.

Key Words: Predicting telecom customer churn, machine learning techniques, Random Forest algorithm, Decision Tree method, XGBoost model, retention of customers, analytics for forecasting, gradient boosting techniques, telecom sector.

1. INTRODUCTION

Predicting telecom churn is an essential activity for telecommunications companies aiming to keep their customers and reduce revenue losses. Customer churn is the term used to describe the situation in which users stop utilizing a telecom service, either by moving to a rival provider or discontinuing their service completely. Understanding the factors affecting the churn and prediction of potential of customer attrition can work for telecom companies take quick actions to improve customer retention. Machine learning techniques have a vital role in identifying data patterns and trends that indicate possible churn, allowing companies to optimize marketing strategies and enhance customer satisfaction. To build an efficient churn prediction model, variety of algorithms can be employed, including Random Forest, Decision Tree, and XGBoost. These algorithms analyze historical customer data, such as call timings, internet uses, billing history, and customer complaints, to identify potential churners. Decision Trees provide an interpretable model for understanding key factors leading to customer churn, while Random Forest, an ensemble method, enhances predictive accuracy by making decision trees. On another side, XGBoost, a powerful gradient-boosting algorithm, refines predictions by optimizing model performance and handling imbalanced datasets effectively. The use of these algorithms in churn prediction enables telecom companies to classify customers into categories of high or low churn risk. By utilizing these forecasts, companies can create focused customer retention approaches like customized promotions, loyalty initiatives, and enhanced customer support. The ability to forecast churn not only reduces customer acquisition costs but also strengthens the long-term relationship between the company and its subscribers. Better strategic planning and decision-making are both made possible by using feature selection approaches to determine which elements are most important in churn. Churn prediction using machine learning is a data-driven approach that helps service providers mitigate customer attrition and improve business sustainability. The combination of Random Forest, Decision Tree, and XGBoost provides strong and precise predictive models that support well-informed business choices. With continuous advancements in machine learning and data analytics, telecom companies can enhance their customer retention strategies, ultimately leading to increased revenue and customer satisfaction.

2. WHAT IS CUSTOMER CHURN?

Client churn, often called customer attrition, occurs when clients stop using a business's goods or services within a certain period of time. In the telecommunications industry, churn occurs when subscribers cancel their service, switch to a competitor, or stop using the network entirely. Churn can be voluntary, where customers leave due to dissatisfaction, pricing, or better offers from competitors, or involuntary, due to factors like payment failures or service terminations.
Forecasting and minimizing customer churn is essential for companies, since customers is generally less expensive than seeking newest ones.

3. LITERATURE REVIEW

1) Md Sayedur Rahman, Md Sahrial Alam, and Md Ikbal Hosen (2022) investigate how effective in forecasting customer churn. The researchers evaluate several models, such as Decision Tree, Random Forest, and boosting methods, to determine which one performs the best. The results indicate that ensemble techniques, especially Random Forest and XGBoost, achieve greater accuracy in

predicting churn by effectively managing intricate relationships within customer data.

2) P. Lalwani, M. K. Mishra, J. S. Chadha, and P. Sethi (2022) describe a system for predicting customer churn using machine learning, which utilizes past customer data to enhance retention strategies. The study evaluates various algorithms and highlights feature selection techniques to enhance model performance. The results demonstrate that hybrid models combining multiple learning techniques can significantly improve churn prediction accuracy.
3) J. K. Sana, M. Z. Abedin, M. S. Rahman, and M. S. Rahman (2022) This research focuses on optimizing churn prediction models through data transformation techniques tailored for the telecommunications industry. By applying feature engineering and preprocessing methods, the study improves model efficiency and accuracy. The findings suggest that well-processed data leads to better customer segmentation and more precise churn forecasts.
4) Prasanth Senthan, RMKT Rathnayaka, Banujan Kuhaneswaran, and BTGS Kumara (2021) present a churn prediction model utilizing XGBoost, showcasing its effectiveness in managing imbalanced datasets and intricate patterns in customer behavior. The research underscores the significance of hyperparameter tuning and data preprocessing in maximizing the performance of the model. The findings indicate that XGBoost surpasses traditional algorithms in forecasting telecom customer churn.
5) VLN RajaGopal Kesiraju and P. Deeplakshmi (2021) This paper proposes a dynamic churn prediction approach using machine learning algorithms, focusing on real-time customer behavior analysis. The study integrates behavioral patterns with predictive analytics to improve retention strategies. The results emphasize that a continuously updating model can better adapt to evolving customer preferences and reduce churn rates.
6) S. Wu, W.-C. Yau, T.-S. Ong, and S.-C. Chong (2021) This research introduces a comprehensive framework that merges churn prediction with customer segmentation, aimed at assisting telecom companies in refining their retention strategies. By grouping customers according to their risk of churning, the model facilitates more precise interventions. The results indicate that using segmentation-based methods improves the efficiency of churn prediction and enhances strategic decision-making.
7) U. F. Gursoy, E. M. Yildiz, M. E. Okay, and M. S. Aktas (2021) This research explores the integration of churn prediction models with promotional strategies in the telecommunication sector. The study evaluates machine learning techniques in identifying at-risk customers and designing personalized promotions to improve retention. The results demonstrate that predictive analytics, when combined with marketing efforts, significantly reduces churn rates.
8) J. Engelmann and S. Lessmann (2021)The authors introduce a novel approach using Conditional Wasserstein GANs to address class imbalance in churn prediction datasets. By generating synthetic customer data, the study enhances model training and improves prediction accuracy. The findings indicate that GAN-based oversampling methods effectively mitigate the challenges of imbalanced learning in churn prediction.
9) Z. Liu, Y. Fang, Y. Liu, and V. W. Zheng (2021)This paper proposes an adversarial graph neural network (GNN) model for churn prediction, leveraging neighbor-anchoring. The study demonstrates that GNNs can capture complex relationships in customer data more effectively than traditional models. The results suggest that deep learning-based approaches offer significant improvements in churn prediction accuracy.
10) L. Almuqren, F. S. Alrayes, and A. I. Cristea (2021) The study explores customer churn behavior prediction using sentiment analysis on Arabic Twitter data. By applying (NLP) and different machine learning techniques, the research identifies early indicators of customer dissatisfaction. The findings highlight the potential of social media mining in augmenting traditional churn prediction models for telecom companies.

4. PROPOSED SYSTEM

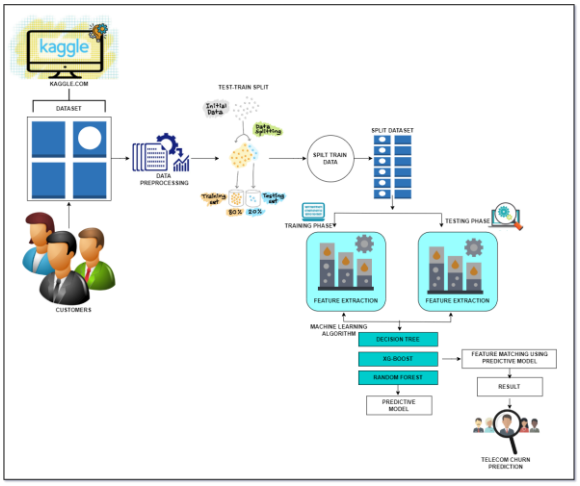

Fig: System Architecture

The telecom churn prediction system shows in the fig. leverages machine learning to anticipate customer attrition. Raw data is sourced from Kaggle and undergoes meticulous preprocessing, encompassing data cleaning, feature selection, and splitting into training and testing subsets The preprocessed data is utilized as input for different machine learning models. Each model is trained separately, and their predictive capabilities are thoroughly assessed. The model that shows the best accuracy is then chosen for deployment. During the operational phase, fresh customer data is consistently input into the deployed model, producing predictions on the likelihood of each customer churning. These predictions empower proactive customer engagement strategies, such as targeted promotions, enhanced customer service, and retention initiatives. Moreover, the system incorporates a feedback loop to dynamically monitor model performance and facilitate retraining as necessary, ensuring ongoing accuracy and effectiveness.

5. ALGORITHM

1) Random Forest :
Random Forest is a popular learning algorithm commonly employed for prediction in the industry because of its effectiveness and reliability. It functions by creating several tree-based models applied to different segments of the dataset and combining their forecasts to enhance the resilience of the model and decrease overfitting. This strategy ensures that the model performs well with new data, making it particularly effective in forecasting customer churn. In the context of telecom churn prediction, Random Forest evaluates different customer attributes, including call duration, data consumption, billing habits, and customer complaints, to assess the probability of churn. By calculating the average predictions from multiple decision trees, it reduces both bias and variance, leading to a more dependable classification model. A high profit of Random Forest is its capability to manage extensive datasets with many variables while yielding high predictive accuracy. It also provides feature importance rankings, allowing telecom companies to identify the most significant factors influencing churn. This insight helps in formulating targeted retention strategies and improving customer satisfaction. Nonetheless, Random Forest can be demanding in terms of computation, particularly when working with large datasets, necessitating substantial processing power for training purposes. Even so, its capability to manage imbalanced datasets and handle missing values makes it a favored option for forecasting customer attrition.

2) Decision Tree :
A Decision Tree is an effective and straightforward machine learning method employed for predicting customer churn in the sector. It operates by systematically dividing the dataset according to feature values, creating a tree-like format where each internal node symbolizes a decision criterion, and the terminal nodes represent classifications of churn or non-churn. Decision Trees are highly interpretable, allowing telecom companies to understand the key factors driving customer attrition. For example, a Decision Tree might reveal that customers with high monthly bills and frequent service complaints are more likely to churn. This level of interpretability makes it easier for business stakeholders to act on the insights derived from the model.To address this, techniques such as pruning and setting depth limits are used to improve the model's generalization. Despite its limitations, Decision Tree remains a valuable tool for initial exploratory analysis and as a foundational component in ensemble methods like Random Forest and XGBoost, which build on its strengths while mitigating its weaknesses.

3) XGBoost for Telecom Churn Prediction
Extreme Gradient Boosting is a highly effective machine learning algorithm that has become well-known for predicting telecom churn because of its speed and accuracy in predictions. This boosting method creates several weak decision trees in succession, with each tree addressing the mistakes made by the earlier one. This step-by-step learning approach improves the model's effectiveness by minimizing both bias and variance, which is why it works so well. for imbalanced datasets commonly found in churn prediction. XGBoost also employs regularization techniques to prevent overfitting, ensuring better generalization to data. One of benefits of XGBoost is its ability to handle missing values and optimize feature selection, making it ideal for complex telecom datasets with diverse customer attributes. The algorithm's speed and scalability allow it to process large datasets efficiently, making it a preferred choice for real-time churn prediction applications. Additionally, XGBoost provides interpretable results through feature importance scores, helping telecom companies identify critical churn factors and devise proactive retention strategies. Despite its computational complexity, the superior performance of XGBoost in churn prediction makes it a highly effective tool for telecom service providers targeting to up the customer retention and lower the churning rates.

6. CONCLUSIONS

Forecasting customer churning rates in the sector using machine learning is a vital strategy for providers to retain their customers and minimize revenue declines. By utilizing algorithms like Random Forest, Decision Tree, and XGBoost, companies can efficiently study customer behavior trends and spot individuals who are likely to churn. Random Forest enhances predictive accuracy through ensemble learning, Decision Tree provides interpretability for understanding key churn factors, and XGBoost optimizes performance through boosting techniques. These models allow telecommunications firms to execute focused retention tactics, including customized deals and proactive interaction with customers. With continuous advancements in machine learning and data analytics, integrating these algorithms into churn prediction implementing robust systems can greatly boost customer loyalty, lower churn rates, and improve overall sustainability for businesses in the competitive telecom sector.

REFERENCES

1. Md Sayedur Rahman, Md Sahrial Alam, and Md Ikbal Hosen (2022) presented a study titled "To Predict Customer Churn Using Different Algorithms" at the International Conference on Decision Aid Sciences and Applications, where they evaluated various machine learning techniques for churn prediction.
2. P. Lalwani, M. K. Mishra, J. S. Chadha, and P. Sethi (2022) explored a machine learning-driven Customer Churn Prediction System in their study published in Computing (Vol. 104, No. 2, pp. 271-294), highlighting effective predictive models for customer retention.

3. J. K. Sana, M. Z. Abedin, M. S. Rahman, and M. S. Rahman (2022) proposed an optimized churn prediction model for the telecom sector, focusing on data transformation techniques to enhance prediction accuracy.
4. Prasanth Senthan, RMKT Rathnayaka, Banujan Kuhaneswaran, and BTGS Kumara (2021) developed a churn prediction model using XGBoost, showcasing its efficiency in handling imbalanced data, as presented at the IEEE International IoT Electronics and Mechatronics Conference (IEMTRONICS).
5. VLN RajaGopal Kesiraju and P. Deeplakshmi (2021) introduced a dynamic churn prediction model that leverages machine learning to analyze customer behavior, as documented in the International Conference on Computer Communication and Informatics (ICCCI).
6. S. Wu, W.-C. Yau, T.-S. Ong, and S.-C. Chong (2021) proposed an integrated framework for churn prediction and customer segmentation in the telecom industry, published in IEEE Access (Vol. 9, pp. 62,118-62,136).
7. U. F. Gursoy, E. M. Yildiz, M. E. Okay, and M. S. Aktas (2021) examined the application of machine learning in churn prediction and promotion strategies for telecom businesses in their study presented at the SAI Intelligent Systems Conference (pp. 276-286).
8. J. Engelmann and S. Lessmann (2021) explored the use of Conditional Wasserstein GANs for oversampling imbalanced tabular data, enhancing churn prediction accuracy, in their work published in Expert Systems with Applications (Vol. 174, Article 114582).
9. Z. Liu, Y. Fang, Y. Liu, and V. W. Zheng (2021) introduced Neighbor-Anchoring Adversarial Graph Neural Networks for churn prediction, leveraging deep learning techniques, in their paper published in IEEE Transactions on Knowledge and Data Engineering.
10. L. Almuqren, F. S. Alrayes, and A. I. Cristea (2021) conducted an empirical study on predicting customer churn behavior using Arabic Twitter data mining techniques, published in Future Internet (Vol. 13, No. 7, Article 175).