ISSN 2277-2685
IJESR/Jan-Mar. 2026/ Vol-16/Issue-1/548-554
R. Srija Reddy *et. al.*, / International Journal of Engineering & Science Research
Full Length Article

Customer Churn Prediction For Telecom

R. Srija Reddy[1], MR. Dr. Kiran B. M[2],CH. Pranitha[3], S. Sai Leela[4], V. Varun Kumar Reddy[5], A. Naveen Kumar[6]
[1,2]Department Of Computer Science And Engineering Sphoorthy Engineering College Hyderabad India
[3,4,5,6]B.Tech Students; Department Of Computer Science And Engineering Sphoorthy Engineering College Hyderabad India
Mail Id; dr.kiran@sphoorthyengg.ac.in[2], radamallasrija@gmail.com[1], chikkelapranitha@gmail.com[3], sheelasaileela@gmail.com[4], varunvemireddy630@gmailoc.om[5], nk999142@gmai.com[6]
Accepted 29-03-2026

ABSTRACT
Customer churn prediction is an important area of interest for businesses in competitive and subscription based business models where customer retention is a key determinant for business success. Accurate prediction of customers who are likely to churn can help businesses design proactive strategies for retaining customers and reducing business losses. In this paper, a design and implementation of a Customer Churn Prediction System using machine learning techniques is presented to accurately predict customers as Churned or Stayed and their probability of Churning The proposed system uses a range of customer data, including demographic characteristics, service usage behavior, billing and contract details, and service engagement. A structured data processing pipeline is designed for accurate data processing and modeling. To overcome common problems in churn prediction models related to class imbalance issues in machine learning models, a hybrid model using SMOTE ENN is implemented for accurate class imbalance handling. Further, a machine learning model using a random forest algorithm is implemented for accurate Predictions. The random forest algorithm is a popular ensemble machine learning technique known for its robustness and high accuracy in dealing with a wide range of data types. The performance of the model can be evaluated using different metrics like accuracy, precision, recall, and F1-score. However, it should be noted that recall should be given prime importance to effectively identify high-risk customers. The proposed system utilizes SQL for storing data in a structured manner and Flask for deploying the application. Interactive visualization has been achieved through Plotly. The proposed system has been designed to effectively cater to the problem of customer churn prediction in an effective manner. The proposed approach utilizes advanced sampling techniques and ensemble learning to effectively predict customer churn, and through the integration of visualization techniques, it can be effectively utilized to take effective decisions regarding retaining customers within an organization.

Index Terms*—Customer Churn, Machine Learning, Predictive Analytics, Data Preprocessing, SMOTE-ENN, Classification, Customer Retention, Business Intelligence, Flask Web Application, Predictive Analytics, Interactive Visualization.*

I. INTRODUCTION

Customer churn prediction is a data-driven analytical technique that helps in the identification of customers who may likely stop doing business with a particular organization in the future. Customer churn prediction is extremely vital in highly competitive business environments such as telecommunication services, e-commerce sites, and online streaming services. The importance of retaining customers in the mentioned business environments can be understood by the fact that studies have proven that it is far much cheaper to retain existing customers than to spend large amounts of funds recruiting new Customers. Customer churn prediction has a significant impact on the profitability of business organizations in that when customers churn, the business organization not only loses the revenue earned from the customers but also has to spend large amounts of funds in the recruitment of new customers to replace the churned ones.Moreover, the prediction of customer churn also helps in the identification of underlying causes of the churn, such as poor services, customer dissatisfaction, and poor customer care services. Customer churn prediction also has a significant impact on the brand reputation of business organizations in that when customers churn, the business organization not only loses the revenue earned from the customers but also has to spend large amounts of funds in the recruitment of new customers to replace the churned ones. Moreover, the prediction of customer churn also helps in the identification of

underlying causes of the churn, such as poor services, customer dissatisfaction, and poor customer care services.This project is concerned with the design and implementation of a Customer Churn Prediction System using state-Of-the-art machine learning algorithms. The proposed system has the potential to effectively process and analyze a given set of structured data related to customers, such as demographic information, usage patterns, contract type, payment type, billing information, tenure, and engagement metrics. By using a given set of historical labeled data, it can effectively identify patterns related to customers and determine the probability of customers churning. The proposed Customer Churn Prediction System is a supervised learning model that utilizes a given set of historical data related to customers and attempts to create a model to predict potential customers who are likely to churn. One of the major challenges faced by a Customer Churn Prediction System is class imbalance, which occurs when a larger percentage of customers are not churning compared to customers who are churning. To solve this challenge, this project proposes implementing a hybrid algorithm known as SMOTE ENN, which can effectively solve class imbalance by creating synthetic data and eliminating noisy data.

The basic prediction model is developed using the Random Forest algorithm, which is an ensemble learning technique that combines multiple decision trees for better accuracy. Random Forest is an appropriate algorithm for the churn prediction problem since it can handle both categorical and numerical variables, offers high accuracy, and can perform feature importance analysis. The complete system is developed as an organized process that includes data extraction, data processing, dealing with class imbalance, feature development, model training, model evaluation, prediction creation, and visualizations. The processed predictions are stored in a SQL database, which is then connected with the Flask-based web application.The visualizations include churn distribution, contract-based churn, revenue impact, and feature importance. These visualizations can help users understand the churn data better. Unlike conventional statistical techniques, this proposed machine learning-based approach can automate churn prediction. This system not only classifies customers as Churned or Stayed but also provides the churn probabilities for the customers. This can help businesses make better decisions than using the conventional classification technique.

II. RELATED WORK

Customer churn prediction has been extensively researched in the areas of data mining, machine learning, and business analytics because of the significant impact it has on customer retention and revenue maximization. Several researchers have employed different statistical and machine learning models to enhance the accuracy of the churn prediction model. This section discusses the significant contributions of researchers in the areas of churn prediction, handling class imbalance problems, and ensemble methods that can be applied to the proposed system. Machine Learning Approaches for Churn Prediction. Customer churn prediction has been widely studied in the fields of data mining, machine learning, and business analytics due to its direct impact on customer retention and revenue optimization. Researchers have explored various statistical and machine learning techniques to improve churn prediction accuracy and interpretability. This section reviews significant contributions in churn prediction, class imbalance handling, and ensemble learning approaches relevant to the proposed system.

A. Machine Learning Approaches for Churn Prediction

The early studies on the prediction of customer churn have mainly employed traditional statistical techniques. In this context, logistic regression has been employed for the purpose of classification. However, the accuracy of the model reduces with the increase in the complexity of the data, which is generally the case with large datasets of customers. Decision trees have also been employed for the classification of customers. However, the accuracy of the model reduces with the increase in the number of data instances, resulting in overfitting. In recent times, with the advent of advanced techniques of machine learning, the accuracy of the model has increased with the use of ensemble techniques such as Random Forest. Random Forest has gained more popularity in recent times with the increase in the accuracy of the model. This is mainly due to the fact that the model reduces the variance of the data with the use of the bagging technique. In this context, it has been identified that the accuracy of the model increases with the use of ensemble learning techniques for the classification of customers with heterogeneous features.

B. Class Imbalance Problem in Churn Datasets

One of the challenges identified in the churn prediction problem is class imbalance. In the real-world dataset, the number of non-churn customers is much larger than the number of churn customers. This can affect the classification model's performance, where the accuracy of the model is high but the number of churn customers identified is very low. To solve this problem of class imbalance, researchers have proposed different resampling techniques. These techniques include random over-sampling, random

under-sampling, and synthetic data generation. One of the most popular techniques for class balancing is the Synthetic Minority Over-sampling Technique (SMOTE). In this technique, artificial data samples are generated for the minority class using the nearest neighbor interpolation technique.

Although the SMOTE technique has proven successful in balancing the dataset, it can also include noise or overlapping data. To solve this problem, researchers have proposed the SMOTE with Edited Nearest Neighbors (SMOTE-ENN) technique. In this technique, not only is the minority class over-sampled, but noise or overlapping data is also removed from the dataset. This technique has proven successful in improving the recall and F1-score of the churn prediction problem.

C. Web-Based Deployment and Visualization

In the past few years, there has been a growing interest in incorporating prediction systems with web-based applications and interactive visualizations. Studies emphasize the need to not only make predictions, but to display the results in a format that is easily interpretable for stakeholders. Interactive visualizations allow decision-makers to easily track changes in customer churn, analyze segmentation risks, and assess revenue implications. Current systems for prediction and analysis incorporate machine learning algorithms with backend frameworks and database systems, allowing for real time predictions and deployment.

III. PROPOSED SYSTEM

The proposed system is intended to be an end-to-end customer churn prediction and analysis system that covers the aspects of data acquisition, data preprocessing, churn detection, risk segmenting, revenue analysis, and business intelligence visualization.

The main idea of the proposed system is to provide support to telecom organizations in the identification of their customers who are likely to churn from their services.

The proposed system works in the form of interconnected components that provide support in the processing of the data related to the customers.

A. System Architecture

In the proposed system architecture, the system is composed of multiple interconnected modules that process the customer information and provide the churn prediction. In the initial stage, the telecom customer information is uploaded into the system using the data acquisition module, which checks the information and saves the data in the system. After the preprocessing stage, the system uses the machine learning module to perform the classification of the data using the Random Forest algorithm with the combination of SMOTE and ENN. Then, the system performs the risk score and segmentation of the customers into low, medium, and high risk. The revenue impact and forecasting module calculates the potential loss of revenue from the churned customers. Finally, the information is represented in the form of interactive dashboards and executive information summaries.

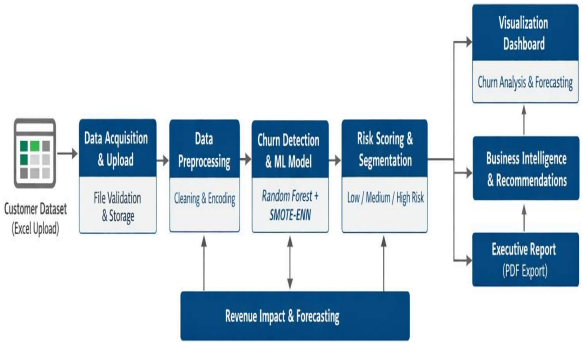

Fig. 1. System Operation Process

B. Data Acquisition and Upload Module

The system also provides the option for the user to upload telecom customer data in the Excel format through the web interface provided by the system. Once the upload is complete, the system will validate the data and store it securely on the server. The system also includes the feature of auto column detection, which helps in detecting important customer information such as Customer Status, Total Revenue, Churn Reason, Contract Type, Internet Type, etc. This module will ensure that the raw customer data is easily integrated into the system.

C. Data Preprocessing Module

Once the dataset is loaded, the data preprocessing module is responsible for preparing the data for further analysis. This module deals with issues such as missing values, inconsistencies, and incorrect data types in the loaded dataset. The column names and formats are also standardized to ensure uniformity with other datasets. Numerical data such as Total Revenue and Monthly Charge is converted to a numeric format, allowing for further computation of the data.

TABLE I: KEY FEATURES USED FOR CHURN PREDICTION

Feature Name	Description
Tenure	Number of months the customer has stayed
Contract Type	Type of service contract
Monthly Charges	Monthly payment made by customer
Internet Type	Type of internet service used
Total Revenue	Total amount paid by the customer

Customer Status	Indicates churned or active customer

D. Churn Detection Module

The churn detection module automatically identifies customers who have ceased using telecom services. This is achieved by using certain parameters, such as the Customer Status column, which may contain values such as Churned, Yes, True, or 1, among others. Using such parameters, the customers are categorized as churned and active. Additionally, the module calculates vital customer Churn related information such as the total number of customers, the number of customers who have churned, and the percentage of customers who have churned, giving a general overview of the situation with regards to customer churn.

E. Risk Scoring and Customer Segmentation Module

The proposed system would provide a churn risk score to each customer depending on various factors such as the duration of the contract, type of contract, monthly bills, and internet services. Depending on these factors, the customers would be classified into Low Risk, Medium Risk, and High Risk categories. This would help telecom companies identify customers who are more likely to churn.

TABLE II: CUSTOMER RISK SEGMENTATION

Risk Level	Description	Suggested Action
Low Risk	Customers unlikely to churn	Maintain engagement
Medium Risk	Customers with moderate churn probability	Offer targeted promotions
High Risk	Customers likely to churn	Immediate retention strategies

F. Revenue Impact and Forecasting Module

The revenue analysis module measures the impact of customer churn on revenue. It calculates the revenue impact by summing up the Total Revenue for churned customers. It also forecasts the possible revenue losses that may occur in the upcoming 6 to 12 months using past customer churn rates. Results are displayed in the form of visuals, such as line charts, making it easier for decision-makers to understand the risks that may occur in the future.

IV. METHODOLOGY

The proposed system for churn prediction will follow the following steps: data collection, data cleaning, feature generation, building the model, and visualization of the insights. The idea is to use the system for precise identification of the customers who are likely to stop their telecom service and provide useful information for improving the retention strategies. The methodology of the proposed customer churn prediction system involves a structured process that includes data collection, preprocessing, feature engineering, model development, and visualization of insights. The goal of this methodology is to accurately identify customers who are likely to discontinue telecom services and provide actionable insights for improving customer retention strategies

A. Data Collection

The whole process begins with the collection of telecom customer information from well-organized data files. The data files contain various information about the customer, their usage of the services provided, the amount they pay, the contract details, and the way they pay their bills. The major factors considered during the analysis include the length of time the customer has been with the company (tenure), the amount the customer pays each month, the amount of revenue generated, the type of internet connection, the type of contract, and the customer's current status. The data is introduced into the system through an Excel-based interface and then saved.

B. Data Preprocessing

Once we have the data, we proceed to the preprocessing stage, which enhances the quality of the information. In the preprocessing stage, we deal with the missing values, inconsistencies, and convert the information into an appropriate format for analysis. For instance, the Total Revenue and the Monthly Charges variables, which are numerical, are converted into the appropriate format. On the other hand, the Contract Type, Internet Service Type, and the Payment Method, which are categorical variables, are encoded. This ensures that the machine learning models operate with reliable information.

C. Handling Class Imbalance

Customer churn datasets are typically imbalanced, where the number of non-churned customers is significantly higher than churned customers. To address this issue, the SMOTEENN technique is applied. The Synthetic Minority Oversampling Technique (SMOTE) generates synthetic samples for the minority class, while Edited Nearest Neighbors (ENN) removes noisy or misclassified instances. This hybrid approach improves the dataset balance and enhances the performance of the prediction model.

D. Feature Selection and Engineering

Feature selection is used to identify the features that most affect customer churn. For example, we analyze the time the customer has been with us, the contract,

the amount the customer pays per month, and the kind of internet service the customer has, among other features, to determine the effect these features have on customer churn. We also use feature engineering to further enhance the features, thus increasing the accuracy of the model in predicting customer churn.

$$Accuracy = \frac{TP}{TP + TN + FP + FN} (1)$$

$$Precision = \frac{TP}{TP + FP} (2)$$

$$Recall = \frac{TP}{TP + FN} (3)$$

$$F1 = 2 \times \frac{Precision + Recall}{Precision + Recall} (4)$$

Where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

Data Visualization and Reporting

The last step in the process is to present the analysis results, which is accomplished through interactive dashboards ,is used to present results such as how customer churn is distributed, customer churn by service type, customer churn by contract type, major reasons for customer churn, and impact on revenue, among others. This will enable telecommunication firms to understand customer churn and support data-driven decision-making.

V. RESULTS AND DISCUSSION

The proposed system, named the Customer Churn Prediction System, will present the results to the end-users in an interactive form using an analytics dashboard. The dashboard will present the key insights to the end-users, such as churn rate, risk segmentation, revenue impact, and the major causes of customer churn. Using these visualizations, the telecom companies will be able to understand the customers' behavior.

A. Customer Churn Overview Dashboard

The dashboard offers a summary of the telecom customer data, providing a summary of key business figures in one place. The figures include the total customers, the number of customers who have churned, the total percentage of customers who have churned, and the revenue at risk. The total number of customers is 6,427, with 2,434 of those customers having churned, giving a percentage of 37.87% of customers who have churned. Additionally, the data indicates that the revenue at risk is approximately 3,588,979.77, indicating the potential loss of revenue due to customer churn.

Fig. 2. Customer Churn Intelligence Dashboard Overview

B. Customer Risk Segmentation

The system we are proposing classifies the customers into Low, Medium, and High Risk based on the churn probability provided by the machine learning model. Such classification enables telecom service providers to identify the customers who have the highest chances of churning the service. From the above visualization, we can see that the majority of the customers belong to the Low Risk group, while a significant number of customers belong to the High Risk group, indicating churn risk for the customers. By focusing on the high-risk group, telecom service providers can launch retention strategies for their customers.

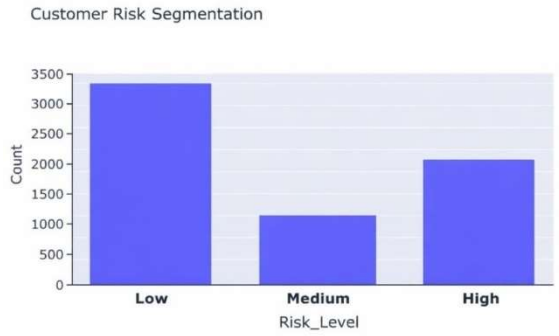

Fig. 3. Customer Risk Segmentation Based on Churn Probability

C. Churn by Contract Type

The system we are proposing classifies the customers into Low, Medium, and High Risk based on the churn probability provided by the machine learning model. Such classification enables telecom service providers to identify the customers who have the highest chances of churning the service. From the above visualization, we can see that the majority of the customers belong to the Low Risk group, while a significant number of customers belong to the High Risk group, indicating churn risk for the customers. By focusing on the high-risk group, telecom service providers can launch retention strategies for their customers.

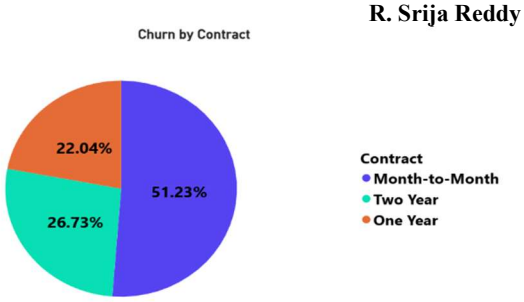

Fig. 4. Customer Churn Distribution by Contract Type

D. Churn by Category

The visualization displays the main buckets of churn reasons, which are Competitor, Dissatisfaction, Attitude, Price, and Other. It is clear that the largest percentage of churn is attributed to the Competitor-related reasons, implying that the main reason why customers are leaving is because the competitor is offering better deals or has a better attitude. However, the fact that dissatisfaction and attitude are also major contributors to churn highlights the fact that telecom service providers need to work on improving the quality of service and the overall customer experience.5. Revenue

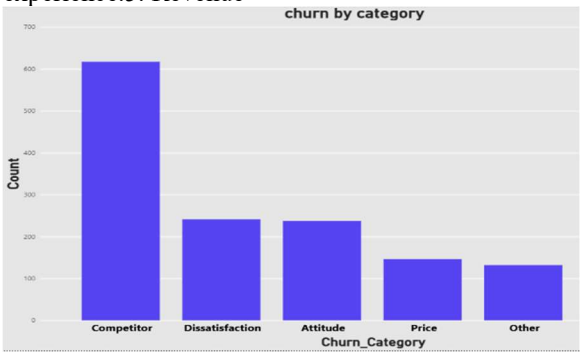

Fig. 5. Customer Churn Distribution by Category

E. Revenue Loss by Churn Reason

The revenue loss analysis essentially shows the impact that various churn reasons have on the organizations revenue. It indicates that churn due to competitor offers, better device options, and better service packages results in the greatest revenue losses. Other factors, such as network reliability issues, poor customer support, and high prices, also contribute to revenue losses. However, the analysis indicates which churn factors have the greatest impact on the organizations revenue, helping the organization identify the areas that need the greatest attention.

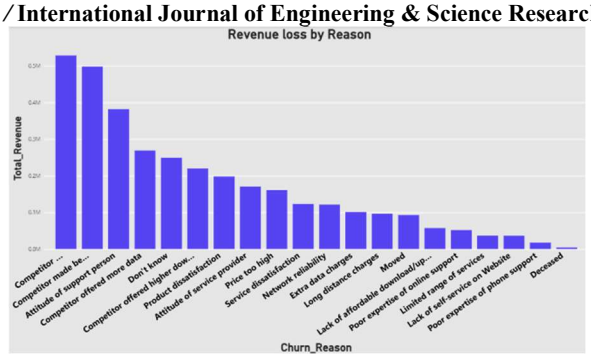

Fig. 6. Revenue Loss Analysis Based on Churn Reasons

F. Feature Importance

Feature importance analysis is useful in determining the key drivers of customer churn prediction models. For the Random Forest model, the importance of each feature is based on the amount it contributes to the prediction process. The key features include total revenue, total charges, monthly charges, age, distance charges, and many more. Understanding the key features helps telecom companies focus on the important factors that influence customer behavior.

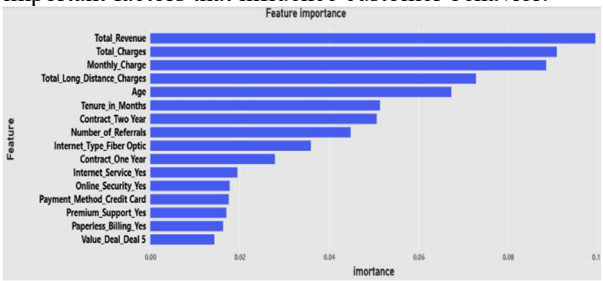

Fig. 7. Feature Importance in Customer Churn Prediction Model

IV FUTURE WORK

In the future, the churn prediction system can be improved by incorporating other sophisticated machine learning algorithms such as Gradient Boosting, XG Boost, etc., into it, which will increase the accuracy level. In addition, it can be extended to support real-time data analysis, which will be helpful for continuously observing customer behavior.

Furthermore, incorporating other interaction data and designing Customer specific strategies will enable telecom companies to reduce churn rates and increase customer satisfaction.

V CONCLUSION

The paper discusses the creation of the Customer Churn Prediction System for the telecommunication industry, which utilizes the concepts of data analytics and machine learning. It uses customer demographic information, service usage, and other relevant information to identify customers who may churn.

Using data preprocessing techniques along with class imbalance handling through the SMOTE-ENN method, the system provides an efficient customer churn prediction.

The interactive dashboards and visualization tools help the telecommunication organizations understand the churn behavior, which in turn helps the organization understand the key factors that cause customer churn.

Therefore, the system helps the organizations take better business decisions that can aid in retaining the customers. Hence, the paper discusses the application of machine learning along with business intelligence for predicting customer churn.

VI REFERENCES

1) S. Jinbo, L. Xiu, and L. Wenhuang, The Application of AdaBoost in Customer Churn Prediction, in ¡em¿Proceedings of the 2007 International Conference on Service Systems and Service Management¡/em¿, 2007.
2) J. Qi and Y. Li, A Novel and Convenient Variable
3) Selection Method for Choosing Effective Input Variables for Telecommunication Customer Churn Prediction Model, in ¡em¿Proceedings of the 2009 IEEE International Conference on Systems, Man and Cybernetics¡/em¿, 2009.
4) L. Liu and H. Ding, Modeling China Telecom Customer Churn Prediction Based on CRISP-DM, in ¡em¿Proceedings of the 2011 International Conference on E-Business and EGovernment (ICEE)¡/em¿, 2011.
5) O. Rezaeian, S. H. R. Shahabi Haghighi, and J. Shahrabi, Customer Churn Prediction Using Data Mining Techniques for an Iranian Payment Application, in ¡em¿Proceedings of the 12th International Conference on Information and Knowledge Technology (IKT)¡/em¿, 2021.
6) A. H. M. Aishwarya, T. Bindhiya, S. Tanisha, B. Soundarya, and C. C. Shanuja, Customer Churn Prediction Using Synthetic Minority Oversampling Technique, in ¡em¿Proceedings of the 4th International Conference on Communication, Computing and Industry 6.0¡/em¿, 2023.
7) P. Pulkundwar, K. Rudani, O. Rane, C. Shah, and S. Virnodkar, A Comparison of Machine Learning Algorithms for Customer Churn Prediction, in ¡em¿Proceedings of the 6th International Conference on Advances in Science and Technology (ICAST)¡/em¿, 2023.