

CUSTOMER CHURN PREDICTION USING LOGISTIC REGRESSION

SK.Harshada Begum¹,G.Jahnavi², V.Harshitha³, Mrs.K.Neeharika⁴,

^{1,2,3}Student, Department of Computer Science & Engineering

Andhra Loyola Institute of Engineering & Technology, Vijayawada, Andhra Pradesh, India

⁴Assistant Professor, Department of Computer Science & Engineering Andhra

Loyola Institute of Engineering & Technology

Andhra Loyola Institute of Engineering & Technology, Vijayawada, Andhra Pradesh, India

Email id: shaikharshada0569@gmail.com, garapatijahnavi@gmail.com,

v.harshitha21064@gmail.com, karra.neeharika@aliet.ac.in

Abstract: This work presents that Customer retention has emerged as a critical strategic priority for subscription based Industries, where the cost of acquiring new customers often triples the cost of retaining existing ones. This project develops a predictive analytics framework to identify potential churners using the Telco Customer Churn dataset. The research follows a comprehensive data science lifecycle, beginning with Exploratory Data Analysis (EDA) to uncover correlations between churn and variables such as contract type, monthly charges, and tenure. Companies can use these findings to create tailored retention strategies like special billing offers and loyalty programs that enhance revenue stability and increase lifetime value. The four different types of Machine Learning models that were developed and evaluated include Logistic Regression, Decision Tree, Random Forest, and XGBoost. The model evaluation process evaluated the ability of each to predict correctly by using Recall and ROC-AUC in addition to Accuracy.

Keywords: Churn of Customer, Life cycle of customer, Logistic Regression

1. INTRODUCTION

Customer churn refers to the situation where customers stop using a company's services or products. In highly competitive industries such as telecommunications, banking, and online services, retaining customers is very important for business growth. The Customer Churn Prediction System is designed to predict whether a customer is likely to leave the service or continue using it. This prediction is done using Machine Learning techniques, specifically the Logistic Regression algorithm. The model analyzes different customer attributes such as tenure, monthly charges, total charges, contract type, and other related features to determine the churn probability.

2. LITERATURE SURVEY

Customer churn prediction has been widely studied using various machine learning techniques to help organizations understand customer behavior and improve retention strategies. Aggarwal et al. (2024) proposed a method for analyzing and predicting churn and retention rates of customers in the telecom industry using the Logistic Regression algorithm. In their approach, a telecom dataset was collected from Kaggle and preprocessing techniques such as data cleaning and feature preparation were applied to improve data quality. The Logistic Regression model was then used to predict whether customers are likely to churn or remain with the service. The performance of the model was evaluated using metrics such as accuracy and precision,

demonstrating the effectiveness of machine learning techniques in predicting customer churn. From the reviewed literature, the following key observations can be made:

- In Existing model no comparison with advanced or hybrid models.
- Limited discussion on real-time deployment and Used only one algorithm.
- Increased computational complexity and Feature selection method may not scale well.
- Requires large datasets for best performance
- Although these methods provide useful insights, many existing systems rely on complex models that require large computational resources and large datasets.

3. PROPOSED SYSTEM

The proposed system aims to develop a Customer Churn Prediction System using a machine learning approach. The system analyzes customer data and predicts whether a customer is likely to churn or continue using the service. The model is developed using the Logistic Regression algorithm, which provides a simple and efficient method for binary classification.

The system uses a dataset containing customer information such as tenure, monthly charges, total charges, contract type, and other service details. Data preprocessing techniques are applied to clean and prepare the dataset before training the machine learning model. After training, the model is used to predict churn for new customer data provided through a user interface.



Fig 1: Proposed System

The system provides a web-based interface where users can enter customer details and receive churn prediction results instantly. This helps organizations understand customer behavior and take preventive measures to retain customers. It Provides accurate prediction of customer churn using machine learning techniques. And uses Logistic Regression which is simple and efficient for classification problems. Helps organizations identify high-risk customers in advance. The proposed system helps businesses make data-driven decisions by identifying potential customer churn in advance. By analyzing customer behavior and predicting churn probability, organizations can implement better retention strategies and improve overall customer satisfaction. This system helps reduce customer loss, increase customer loyalty, and improve business performance

4. METHODOLOGY

The methodology of the Customer Churn Prediction System involves several steps starting from data collection to prediction

1. Data Collection:

The dataset used in this project contains customer information such as tenure, monthly charges, total charges, contract type, and other service-related attributes.

2. Data Preprocessing:

In this step, missing values are handled, duplicate records are removed, and data is converted into a consistent format.

3. Feature Selection:

Not all features in the dataset are useful for prediction, so irrelevant or less important features are removed.

4. Model Training:

The Logistic Regression algorithm is used in this project to train the model. During training, the model learns from historical data and builds a predictive model.

5. Model Evaluation:

This step helps in determining whether the model is reliable or needs improvement. If the performance is not satisfactory, the model can be refined and retrained..

6. Prediction :

The user enters customer details through the web interface, and the system processes the input using the trained model. The model then predicts whether the customer is likely to churn or not. The result is displayed to the user, helping organizations take appropriate actions to retain customers.

5. PROPOSED SYSTEM RESULTS

The proposed system uses the Logistic Regression algorithm, which is a widely used machine learning technique for classification problems. Logistic Regression analyzes customer attributes and predicts whether a customer belongs to the churn or non-churn category. The algorithm learns patterns from historical customer data and uses this knowledge to make predictions on new customer records. This approach allows organizations to identify customers who are at risk of leaving the service and take proactive measures to retain them.

- Users can enter customer details through input fields provided on the web page. The system processes this information and sends it to the machine learning model for prediction. The prediction result is then displayed on the screen, indicating whether the customer is likely to churn or not. This simple interface makes the system easy to use for non-technical users.
- Once the user provides the required customer details, the system analyzes the data and generates a prediction. Based on the prediction result, the system may also provide recommendations to help organizations take suitable actions.
- For example, if the model predicts a high risk of churn, companies may offer special discounts, improve customer support, or introduce loyalty programs to retain the customer.



Fig 2. Our Design Dashboard

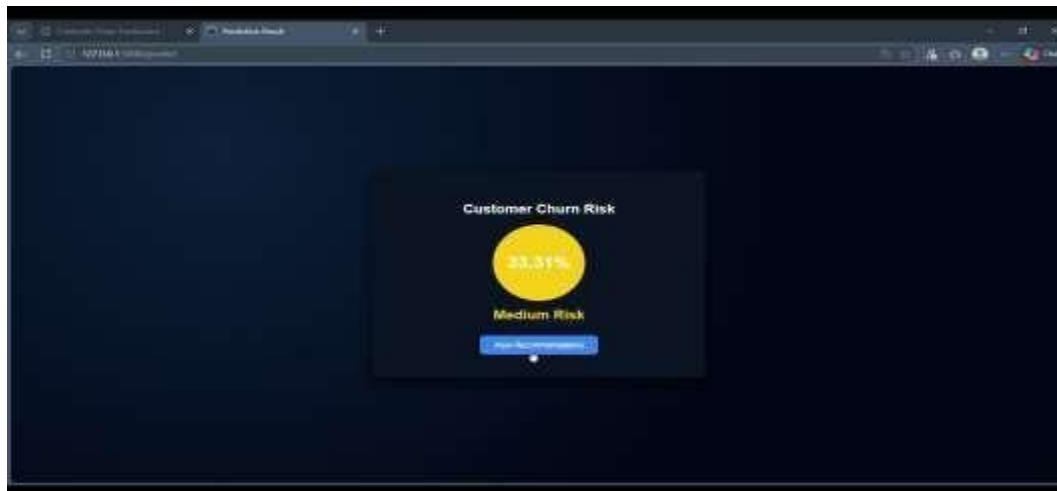


Fig 3. Prediction Risk Percentage

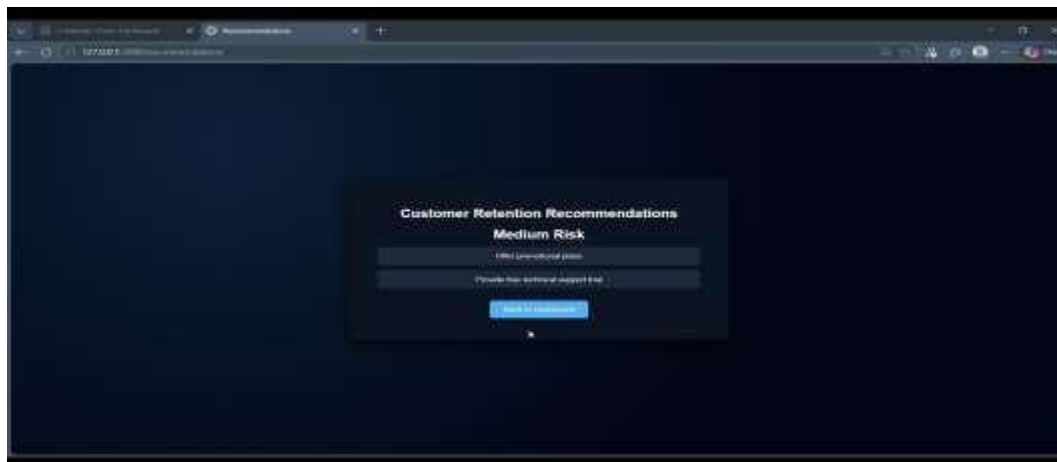


Fig 4. Recommendations for Retention

6. CONCLUSION

In this paper, a Customer Churn Prediction System is developed using machine learning techniques to analyze customer behavior and predict whether a customer is likely to leave a service. The system uses the Logistic Regression algorithm to classify customers into churn and non-churn categories based on various attributes such as tenure, monthly charges, total charges, and contract type. The system successfully processes customer data, performs preprocessing, trains the machine learning model, and generates accurate predictions. A user-friendly web interface is developed using HTML and Flask, allowing users to enter customer details and obtain prediction results easily. The integration of frontend and backend ensures smooth communication and real-time prediction.

7. REFERENCES

- [1] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, 3rd ed. Hoboken, NJ, USA: Wiley, 2013.
- [2] J. Hadden, A. Tiwari, R. Roy, and D. Ruta, "Churn prediction: Does technology matter?," *Int. J. Intell. Technol.*, vol. 1, no. 2, pp. 104–110, 2006.
- [3] Google, "Rules of Machine Learning: Best Practices for ML Engineering," [Online]. Available: <https://developers.google.com/machine-learning/guides/rules-of-ml>
- [4] IBM Sample Data, "Telco Customer Churn," Kaggle Dataset. [Online]. Available: <https://www.kaggle.com/blastchar/telco-customer-churn>
- [5] Pallets, "Flask Documentation." [Online]. Available: <https://flask.palletsprojects.com/>
- [6] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011