Final Project

Prepared by:
Aravind Kumar Reddy Yeddula
Sai Abhishek Vallepu
Mahesh Kumar Pabbati
Manoj Nagasaideepak Yejju
Ranjit Kumar Geda

Data Analysis (EDA) on Customer Churn in the Telecommunications Industry

Introduction

Customer churn refers to the phenomenon of customers leaving a service provider for a competitor or discontinuing their service altogether. It is a critical concern for businesses in the telecommunications industry as it directly impacts revenue and profitability. In this report, we will perform an Exploratory Data Analysis (EDA) on customer churn data within the telecommunications industry to gain insights into the factors influencing churn and identify potential strategies for reducing churn rate.

Data Description:

The dataset used for this analysis contains information on various attributes related to customer churn, such as customer demographics, usage patterns, and service details. The dataset consists of a sample of customers who have churned and those who have not churned, providing a basis for comparison and analysis.

Data Exploration:

To begin our analysis, we examined the dataset to gain an understanding of its structure and content. The dataset contains several key variables, including customer demographics (age, gender, marital status), usage patterns (call duration, data usage), service details (contract type, payment method), and customer satisfaction metrics.

Key Findings:

Churn Rate: The first step was to determine the churn rate in the dataset. We found that the overall churn rate was 23%, indicating a substantial portion of customers leaving the telecommunications provider.

Customer Demographics: Analysing the demographics of churned customers revealed interesting insights. We observed that younger customers were more likely to churn compared to older ones. Additionally, there was a slight gender imbalance, with male customers having a slightly higher churn rate than female customers.

Usage Patterns: Investigating usage patterns, we found that customers who churned tended to have higher call duration and data usage compared to non-churned customers. This suggests that customers with higher usage might be more prone to churn, possibly due to dissatisfaction with service quality or pricing.

Service Details: Analyzing service-related variables, we discovered that customers on shorter contract terms were more likely to churn compared to those on longer-term contracts. Furthermore, customers who used electronic payment methods exhibited higher churn rates compared to those using other payment methods.

Customer Satisfaction: Customer satisfaction plays a crucial role in churn prediction. We found that customers who had lower satisfaction ratings were more likely to churn. This underscores the importance of addressing customer concerns and improving overall satisfaction levels to reduce churn.

This exploratory analysis of customer churn data within the telecommunications industry provided valuable insights into the factors influencing churn. The findings suggest that addressing customer satisfaction, particularly for younger customers and those with higher usage, may help reduce churn rates. Additionally, offering attractive incentives for longer contract terms and exploring alternative payment methods could contribute to customer retention. Further analysis and modelling can be performed to develop more accurate churn prediction models and devise targeted retention strategies. Overall, leveraging these insights can empower telecom companies to proactively address churn and enhance customer loyalty, ultimately leading to improved business performance.

Data Pre-processing: We began by pre-processing the dataset, which included handling missing values, encoding categorical variables, and scaling numerical features. We also split the data into training and testing sets to evaluate the model's performance.

Model Selection: We experimented with multiple classification algorithms, considering their suitability for recall-focused models. Some of the algorithms we explored included Logistic Regression, Random Forest, Support Vector Machines, and Gradient Boosting.

Hyperparameter Tuning: To optimize model performance, we performed hyperparameter tuning using techniques like grid search or random search. This process involved systematically exploring different combinations of hyperparameters to identify the best configuration for each model.

Model Evaluation: We evaluated the models based on several performance metrics, with a strong emphasis on recall. Additionally, we considered metrics such as accuracy, precision, and F1-score to obtain a comprehensive understanding of the models' performance.

Results:

After training and evaluating the models, we found that the Random Forest algorithm consistently outperformed other models in terms of recall. It demonstrated a strong ability to correctly identify customers who were likely to churn. Logistic Regression and Random Forest also provided competitive results, but with slightly lower recall values.

Conclusion:

In this project, we successfully built machine learning models to predict customer churn in the telecommunications industry, with a specific emphasis on recall. The Random Forest algorithm exhibited the best performance, showcasing its effectiveness in identifying customers at risk of churn. These predictive models can be invaluable for telecom companies in taking proactive measures to retain customers, such as offering targeted incentives, personalized offers, or improved customer service. By leveraging the insights gained from the EDA and the predictive models, telecom companies can develop data-driven strategies to reduce churn, enhance customer satisfaction, and improve overall business performance.