Customer Churn Prediction

Milestone: Project Proposal

Group 03
Pramoth Guhan
Harith Anand

(857) 891-6677 (Tel of Student 1)
(857) 423-5587 (Tel of Student 2)

guhan.p@northeastern.edu
anand.ha@northeastern.edu

Percentage of Effort Contributed by Student 1: 50

Percentage of Effort Contributed by Student 2: 50

Signature of Student 1: Pramoth Guhan

Signature of Student 2: Haritha Anand

Submission Date: 03/15/2024

Problem Setting:

The telecommunications industry is highly competitive, with companies constantly seeking ways to reduce customer churn – the rate at which customers discontinue their service. High churn rates can significantly impact a company's revenue and growth prospects. By understanding the factors why customers churn and predicting which ones are at risk, companies can create targeted retention strategies. This project aims to analyze customer data to identify key predictors of churn and develop a predictive model to forecast customer churn. The challenges include dealing with large and complex datasets, ensuring the accuracy of the predictive model, and identifying actionable insights from the analysis.

Problem Definition:

The specific problem this project addresses is the prediction of customer churn in the telecommunications industry. By analyzing customer data, we aim to answer the following questions.

1. What are the key factors that contribute to customer churn?
2. Can we predict which customers are likely to churn in the near future?
3. What strategies can be implemented to reduce churn based on the insights gained from the data?

Data Sources:

The primary data source for this project is the "Telco Customer Churn" dataset available on Kaggle [1]. This dataset includes various customer attributes and their churn status. The dataset was originally provided by IBM as part of their sample data sets to help predict customer churn in the telecommunications industry.

Data Description:

The "Telco Customer Churn" dataset comprises several columns representing customer attributes and their churn status. Key attributes include:

- CustomerID: Unique identifier for each customer
- Gender: Customer's gender (Male/Female)
- SeniorCitizen: Indicates if the customer is a senior citizen (1) or not (0)

- Tenure: Number of months the customer has been with the company
- PhoneService: Indicates if the customer has phone service (Yes/No)
- InternetService: Type of internet service (DSL, Fiber optic, No)
- MonthlyCharges: The amount charged to the customer monthly
- TotalCharges: The total amount charged to the customer

The dataset contains 7043 rows and 21 columns, providing a comprehensive view of customer attributes and their churn status. This data can be used to explore and develop predictive models to identify key factors influencing customer churn in the telecommunications industry.

Project Planning:

Key Planned Milestones

1. Data Preprocessing: Clean the dataset by handling missing values, encoding categorical variables, and normalizing numerical variables. This will be done by Haritha Anand.
2. Exploratory Data Analysis (EDA): Analyze the dataset to understand the distribution of key variables and identify patterns related to customer churn. This will also be handled by Haritha Anand.
3. Feature Selection: Identify the most relevant features that contribute to customer churn using statistical tests and feature important techniques. This task will be carried out by Pramoth Guhan.
4. Model Development: Develop and train machine learning models to predict customer churn. Evaluate several models (e.g., logistic regression, decision trees, random forests) to select the best performer. This will be done by Pramoth Guhan.
5. Model Evaluation: Evaluate the model's performance using metrics such as accuracy, precision, recall, and F1-score. Perform cross-validation to ensure the model's generalizability. This responsibility also lies with Pramoth Guhan.
6. Insight Generation and Recommendations: Based on the model's findings, generate insights into the key factors driving churn and recommend strategies to reduce churn. This final step will be a collaborative effort between Haritha Anand and Pramoth Guhan.

Challenges

Data Quality: Ensuring the dataset is clean and free of errors that could affect the analysis.

Feature Selection: Identifying the most relevant features from a large set of variables can be challenging.
Model Complexity: Developing a model that is both accurate and interpretable.
Implementation of Findings: Translating the insights from the data analysis into actionable business strategies.

This project plan provides a structured approach to addressing the problem of customer churn in the telecommunications industry, leveraging data analytics to generate actionable insights.

Citations:
[1] https://www.kaggle.com/datasets/blastchar/telco-customer-churn?resource=download