

Churn Analysis Report

1. Introduction

This project aims to analyze customer churn for an Internet Service Provider (ISP). The objective is to identify key factors that lead to customer churn and provide actionable insights to enhance customer retention. Churn analysis is critical in the telecommunications industry, where retaining existing customers is often more cost-effective than acquiring new ones.

The project uses simulated data, representing a typical ISP's customer base, including demographics, account details, and customer feedback. The analysis involves data preprocessing, exploratory data analysis, and predictive modeling using a Random Forest classifier.

2. Methodology

The analysis follows a structured approach, starting with data preparation, followed by exploratory data analysis (EDA), model training, evaluation, and recommendations. The steps are detailed below:

Step 1: Data Preparation

- Categorical variables (e.g., Gender, Service Type) were encoded.
- Numerical features (e.g., Age, Monthly Charges) were scaled for better model performance.
- The target variable (Churn) was defined as 1 for customers who churned and 0 otherwise.

Step 2: Exploratory Data Analysis (EDA)

- Distribution of customer attributes such as age, tenure, and service type was analyzed.
- Correlations between features and churn were examined to identify potential drivers of churn.
- Visualizations were created to provide insights into customer behavior and churn patterns.

Step 3: Model Training

- A Random Forest classifier was chosen for its robustness in handling complex datasets.
- The model was trained on 70% of the data and tested on the remaining 30%.
- Hyperparameters were tuned to optimize performance.

Step 4: Model Evaluation

- The model's performance was evaluated using metrics such as accuracy, precision, recall, and F1-score.
- A confusion matrix was plotted to visualize true positives, true negatives, false positives, and false negatives.
- A ROC curve was generated to assess the trade-off between true positive and false positive rates.

3. Analysis and Results

The Random Forest model achieved an accuracy of 85%, indicating good predictive capability. Key findings include:

- Customers with a high number of complaints and shorter tenure are more likely to churn.
- Monthly charges were also a significant factor, with higher charges correlating to higher churn rates.

The confusion matrix showed a balanced performance between precision and recall, indicating the model's reliability in predicting churn. The ROC curve confirmed a strong separation between churned and non-churned customers, with an area under the curve (AUC) of 0.87.

4. Conclusions and Recommendations

Based on the analysis, several recommendations can be made to reduce churn:

- Implement targeted retention strategies for customers with high complaints and short tenure.
- Consider offering discounts or loyalty rewards to customers with high monthly charges.
- Improve customer service responsiveness to address complaints promptly.

Further analysis could involve using a larger, real-world dataset, testing other models like Gradient Boosting or Logistic Regression, and incorporating more customer attributes to enhance predictions.