Telecom Churn Prediction Analysis

This document presents the analysis done for Telecom Churn Prediction using customer-level data. The main steps include data preprocessing, feature engineering, handling missing values, tagging churn, training a Logistic Regression model, and identifying the top features contributing to churn prediction.

Step 1: Data Preprocessing

In this step, we loaded a dataset containing telecom customer data. Basic data checks were performed to understand the shape of the dataset and the missing values.

Shape of dataset: (5000 rows, X columns)

Missing values (top 10 columns with most missing values):

[Missing value summary]

A heatmap visualizing the missing values was also generated to provide a clearer view.

Step 3: Filter High-Value Customers

High-value customers were defined based on the total recharge amount in months 6 and 7. Customers above the 70th percentile of recharge amounts were considered high-value.

The filtered data includes the top 30% of customers who have higher average recharge amounts in these months.

Step 4: Churn Tagging

Churn was tagged based on the customer activity in month 9. If there was no activity in calls, outgoing calls, or data usage, the customer was labeled as churn (1).

The churn tagging process was based on the following conditions:

- No incoming or outgoing calls
- No data usage (both 2G and 3G)

The final dataset includes the churn labels for high-value customers.

Step 5: Feature Engineering

In this step, non-numeric columns were dropped, and missing values were imputed with the median of the respective columns.

After feature engineering, the dataset was cleaned, ready for modeling.

Step 6: Logistic Regression Model

The dataset was split into training and testing sets, and the features were standardized using StandardScaler. A Logistic Regression model was trained on the data.

Model performance was evaluated using the following metrics:

- Classification report (precision, recall, f1-score)
- ROC-AUC score

[Include classification report here]

ROC-AUC: [Add ROC-AUC value here]

Step 8: Feature Importance

The top 10 important features for predicting churn were extracted based on the coefficients of the Logistic Regression model.

A bar chart was generated to visualize these important features.