



# Telecom Churn Prediction and Analysis

**Created By:** Faisal Hakimi

**Date:** May 10, 2024

## Abstract

This project explores and analyzes a Telecom Churn dataset to understand factors contributing to customer churn and to develop a predictive model for forecasting customer churn. The analysis includes data preprocessing, exploratory data analysis (EDA), feature engineering, predictive modeling using various algorithms, model tuning, interpretation of results, and recommendations for reducing churn.

# Telecom Churn Prediction and Analysis

## 1. Introduction

Customer churn is a significant challenge for telecom companies, leading to revenue loss and impacting customer acquisition costs. This project aimed to explore and analyze factors contributing to customer churn in a telecom company's dataset. The goal was to develop a predictive model that identifies customers at high risk of churn with high accuracy. Techniques from Python's data manipulation, visualization, and predictive modeling libraries were employed.

## 2. Data Description

The Telecom Churn dataset comprised information on customers, including:

- CustomerID (unique identifier)
- Gender (Male, Female)
- Age (customer age)
- Tenure (months with the company)
- ServiceCalls (number of customer service calls)
- MonthlyCharges (monthly bill amount)
- TotalCharges (total amount charged)
- Churn (customer churn status - Yes or No)

## 3. Methodology

The project followed these steps:

- **Data Preprocessing:** Missing values were handled, data types converted, and necessary transformations performed.
- **Exploratory Data Analysis (EDA):** The distribution of key variables and their relationships, especially with churn, were analyzed. Insights were visualized with charts and plots.
- **Feature Engineering:** New features potentially impacting churn were derived, and relevant features were selected for model building.
- **Predictive Modeling:** Data was split into training and testing sets. Multiple models were chosen for churn prediction (Logistic Regression, Random Forest, SVM, AdaBoost, XGBoost, K-Nearest Neighbors). Models were trained and evaluated with accuracy, precision, recall, F1-score, and ROC-AUC metrics.
- **Model Tuning:** Model parameters were optimized through cross-validation and grid search techniques to improve performance.
- **Interpretation and Conclusion:** Model results were interpreted to identify key churn predictors. Actionable insights and recommendations for churn reduction were provided.
- **Presentation:** A presentation summarizing findings was created for a technical audience with clear visualizations and explanations.

## 4. Results and Discussion

### Data Preprocessing:

- Missing values for Total Charges (11 instances) were identified and removed due to the small number.
- Exploratory Data Analysis (EDA):

**Customer Churn:** The analysis revealed an average churn rate aligning with industry trends (around 1.9% - 2%).

### Contract Length and Churn:

Month-to-month contracts exhibited a positive correlation with churn, indicating customers are more likely to switch providers with this flexibility.

Two-year contracts showed a negative correlation with churn, suggesting these contracts lock customers in and reduce churn.

**Services and Churn:** Interestingly, services like online security, streaming TV, online backup, and tech support (without internet connection) had a negative correlation with churn. This suggests customers who value these add-on services are more likely to remain with the company.

### Demographics:

The customer base was relatively balanced by gender (around 50% male and 50% female).

A significant portion of the customers were younger people, with only 16% being senior citizens.

Roughly 50% of customers had partners, and 30% had dependents. There was no significant difference in these distributions by gender or senior citizen status.

### Customer Account Information:

Tenure distribution showed a concentration of customers with short tenures (1 month) and longer tenures (around 72 months), potentially reflecting different contract lengths.

The majority of customers (74%) did not churn, indicating a class imbalance in the data (more non-churning customers).

## Conclusion

This project successfully analyzed a telecom churn dataset to develop a churn prediction model. Key factors influencing churn were identified, including contract length and value-added services. The findings provide valuable insights for telecom companies to develop targeted retention strategies and reduce customer churn.