Customer Churn Analysis for Telecom Industry

Introduction

Customer churn - when users stop using a service - is a critical problem in the telecom industry, directly impacting revenue and growth. With growing competition and low switching costs, telecom companies need to proactively identify and retain at-risk customers.

This project focuses on analyzing customer behavior, uncovering churn patterns, and building a data-driven approach to support business decisions using machine learning and Power BI.

Abstract

The goal of this project is to perform an end-to-end churn analysis using a real-world telecom dataset. It includes:

- Data preprocessing and exploratory data analysis (EDA)
- Machine learning model development to classify customers likely to churn
- Customer segmentation based on churn probability
- Interactive dashboard creation using Power BI for stakeholder insights

By identifying the key drivers behind customer churn, this project provides meaningful insights and retention strategies to support business goals.

Tools Used

- Python: Data cleaning, feature engineering, EDA, model building
  - Libraries: Pandas, Scikit-learn, Matplotlib, Seaborn
- Power BI: Dashboarding and visual reporting
- Jupyter Notebook: Code development and experimentation
- GitHub: Version control and project hosting

Steps Involved in Building the Project

1. Data Collection
   - Used the Telco Customer Churn dataset from Kaggle.

Customer Churn Analysis for Telecom Industry

2. Data Preprocessing

- Handled missing values and inconsistent data
- Converted categorical variables into numerical format (label encoding/dummies)
- Performed feature scaling

3. Exploratory Data Analysis (EDA)

- Analyzed distributions, relationships, and class imbalance
- Identified trends related to churn (e.g., complaints, tenure, recharge gaps)

4. Model Building

- Applied a Random Forest Classifier to predict churn
- Evaluated performance using accuracy, confusion matrix, and classification report

5. Customer Segmentation

- Classified customers into:
  - At-Risk (High probability of churn)
  - Loyal (Low probability of churn)
  - Dormant (Inactive or low engagement)

6. Power BI Dashboard

- Built interactive visuals to highlight:
  - Churn distribution
  - Demographic and behavioral trends
  - KPIs (average tenure, complaints, call wait time)
- Designed for non-technical stakeholders to explore data and insights

Conclusion

This project successfully identifies key drivers of customer churn in a telecom business using machine learning and visual storytelling. The combination of Random Forest modeling and an interactive Power BI dashboard enables