Customer Churn Analysis for a Telecommunications Company

---

1. Introduction

Project Overview

This project aims to analyze customer churn data from a telecommunications company to identify patterns and factors contributing to customer attrition. By understanding these factors, the company can develop strategies to retain customers and reduce churn.

Objectives

- Efficiently predict whether a customer will churn.
- Understand customer behavior, focusing on what keeps customers using the services and what causes them to leave.

---

2. Data Source

Dataset

- **Source**: Telco Customer Churn dataset from Kaggle.
- **Attributes**:
  - **Identifier**: Unique ID number of the customer (e.g., customerID).
  - **Target Variable**: Churn status (whether the customer churned or not).
  - **Demographic Information**: Gender, Senior Citizen status, Partner status, Dependents.
  - **Customer Account Information**: Tenure, Contract type, Paperless Billing, Payment Method, Monthly Charges, Total Charges.
  - **Services Information**: Phone Service, Multiple Lines, Internet Service, Online Security, Online Backup, Device Protection, Tech Support, Streaming TV, Streaming Movies.

Company Goals

- Reduce customer churn by identifying key drivers.
- Implement strategies to retain customers based on the analysis.

3. Data Collection

Tools Used

- **Programming Language**: Python.
- **Libraries**: pandas for data manipulation.

Process

1. Download the Telco Customer Churn dataset from Kaggle.
2. Load the dataset into a Jupyter Notebook using pandas.

4. Data Exploration

Initial Exploration

- **Objective**: Understand the dataset's structure and contents.
- **Steps**:
    - Display basic information such as column names, data types, and missing values.
    - Generate summary statistics to get an overview of data distribution.

5. Data Cleaning

Steps Involved

- **Handling Missing Values**: Impute missing values or remove affected rows/columns.
- **Converting Categorical Variables**: Use techniques like one-hot encoding or label encoding.
- **Standardizing Features**: Normalize numerical features if necessary.

6. Exploratory Data Analysis (EDA)

Visualization

- **Target Variable**: Visualize the distribution of churn using a bar plot.

- **Feature Relationships**: Explore the relationships between features and churn using bar plots, box plots, and heatmaps.

Correlation Analysis

- Identify key features influencing churn through correlation analysis and feature importance scores.

---

7. Feature Engineering

Objective

- Enhance the dataset with new features derived from existing data.

Examples

- **Total Services Used**: Calculate the total number of services a customer uses.
- **Customer Tenure**: Use domain knowledge to derive meaningful features like customer tenure.

---

8. Model Building

Model Selection

- **Algorithms**: Logistic Regression, Decision Trees, Random Forests.
- **Data Split**: Divide the dataset into training and testing sets.

Evaluation Metrics

- **Metrics Used**: Accuracy, Precision, Recall, F1-score, ROC-AUC (Receiver Operating Characteristic - Area Under the Curve).

---

9. Model Evaluation

Hyperparameter Tuning

- Use GridSearchCV or RandomizedSearchCV to optimize model performance.

Final Model Evaluation

- Compare the final model's performance with the baseline model using ROC-AUC and other relevant metrics.

---

10. Results and Interpretation

Key Findings

- Interpret the analysis results to identify the primary factors driving customer churn.

Visualizations

- Include feature importance plots, confusion matrices, and ROC curves to communicate findings.

Actionable Insights

- Provide actionable recommendations for reducing customer churn based on the analysis.

---

11. Conclusion and Recommendations

Summary

- Recap the key findings and the importance of these insights for the telecommunications company.

Recommendations

- Suggest specific actions the company can take to improve customer retention.

---

13. Tools and Technologies

Programming Language

- **Python**

Libraries

- pandas, numpy, matplotlib, seaborn, scikit-learn, xgboost

IDE

- Jupyter Notebook

---

14. Project Timeline

Week 1: Data Collection and Exploration

- Download and explore the dataset.

Week 2: Data Cleaning and EDA

- Clean the data and perform exploratory analysis.

Week 3: Feature Engineering and Model Building

- Create new features and build the machine learning model.

Week 4: Model Evaluation and Optimization

- Evaluate the model and perform hyperparameter tuning.

Week 5: Results Interpretation and Documentation

- Interpret the results and document the findings.