

Telco Customer Churn Analysis

Overview

This project analyzes customer data from a telecommunications company to predict customer churn. The analysis involves data loading, cleaning, feature engineering, and applying machine learning models. The goal is to identify factors that contribute to customer churn and develop a predictive model.

Data Source

The data is from a telecommunications company and contains information about customer demographics, services used, and account details.

Project Structure

The project is organized in a Jupyter Notebook (Telco Customer Churn.ipynb) and follows these steps:

1. Data Loading and Exploration:
 - Loads the dataset from a CSV file.
 - Displays the first few rows and checks the data structure.
 - Examines data types and missing values.
2. Data Cleaning and Preprocessing:
 - Converts the TotalCharges column to numeric, handling errors.
 - Fills missing values in TotalCharges with 0.
3. Feature Engineering:
 - Creates dummy variables for categorical features.
 - Splits the data into training and testing sets.
4. Model Training and Evaluation:
 - Trains and evaluates a Logistic Regression model.
 - Trains and evaluates a Random Forest Classifier model.
 - Calculates accuracy, F1 score, and generates a classification report.
 - Visualizes the distribution of predicted churn probabilities.

Tools and Libraries Used

- Python: Primary programming language.
- Pandas: For data manipulation and analysis.
- Scikit-learn: For machine learning (Logistic Regression, Random Forest), data preprocessing, and model evaluation.
- Seaborn: For data visualization.
- Matplotlib: For data visualization.

Dataset Description

The dataset includes the following columns:

- **customerID**: Unique identifier for each customer.
- **gender**: Customer's gender.
- **SeniorCitizen**: Indicates if the customer is a senior citizen (0 or 1).
- **Partner**: Indicates if the customer has a partner (Yes or No).
- **Dependents**: Indicates if the customer has dependents (Yes or No).
- **tenure**: Number of months the customer has been with the company.
- **PhoneService**: Indicates if the customer has phone service (Yes or No).
- **MultipleLines**: Indicates if the customer has multiple phone lines (Yes, No, or No phone service).
- **InternetService**: Customer's internet service provider (DSL, Fiber optic, or No).
- **OnlineSecurity**: Indicates if the customer has online security (Yes, No, or No internet service).
- **OnlineBackup**: Indicates if the customer has online backup (Yes, No, or No internet service).
- **DeviceProtection**: Indicates if the customer has device protection (Yes, No, or No internet service).
- **TechSupport**: Indicates if the customer has technical support (Yes, No, or No internet service).
- **StreamingTV**: Indicates if the customer has streaming TV (Yes, No, or No internet service).
- **StreamingMovies**: Indicates if the customer has streaming movies (Yes, No, or No internet service).
- **Contract**: The type of contract the customer has (Month-to-month, One year, or Two year).
- **PaperlessBilling**: Indicates if the customer has paperless billing (Yes or No).
- **PaymentMethod**: The customer's payment method (e.g., Electronic check, Mailed check).
- **MonthlyCharges**: The customer's monthly charges.
- **TotalCharges**: The customer's total charges.
- **Churn**: Indicates if the customer churned (Yes or No).

How to Run the Project

1. Ensure you have Python installed (preferably Python 3.6 or higher).
2. Install the required Python libraries by running:
3. `pip install pandas numpy scikit-learn seaborn matplotlib`
4. Open the Jupyter Notebook (Telco Customer Churn.ipynb) and run each cell sequentially to reproduce the analysis.

Conclusion

This project analyzes telecom customer data to predict churn. It provides insights into the factors that influence churn and develops a machine learning model that can be used to identify customers at risk of leaving.