

Customer Churn Prediction Using Machine Learning

Project Description

Introduction

Customer churn is a major challenge for businesses, as losing customers directly impacts revenue. This project focuses on predicting customer churn using machine learning techniques. By analyzing customer data, businesses can proactively identify at-risk customers and take necessary actions to improve retention.

Objectives

- Develop a machine learning model to predict whether a customer will churn.
 - Perform **Exploratory Data Analysis (EDA)** to understand key factors influencing churn.
 - Preprocess data by handling missing values, encoding categorical variables, and standardizing numerical features.
 - Train and evaluate a **Random Forest Classifier** for churn prediction.
 - Visualize model performance using confusion matrices, ROC curves, and feature importance plots.
-

Dataset

- The dataset used in this project comes from a **Telco Customer Churn dataset**.
- It contains customer demographics, service usage details, contract information, payment methods, and a churn indicator.

Key Features:

- **Demographic Information:** Gender, Senior Citizen, Partner, Dependents.
 - **Service Usage:** Internet Service, Streaming, Tech Support, Tenure.
 - **Contract and Billing Details:** Payment Method, Monthly Charges, Total Charges, Contract Type.
 - **Target Variable: Churn (Yes/No)**
-

Exploratory Data Analysis (EDA)

- Analyzed churn rates across different customer segments.
- Visualized distributions of numerical features like tenure and monthly charges.

- Explored correlations between churn and categorical variables using bar plots.
-

Data Preprocessing & Feature Engineering

- Removed unnecessary columns (**Customer ID**).
 - Converted **TotalCharges** to numeric and handled missing values.
 - Encoded categorical variables using **Label Encoding**.
 - Scaled numerical features using **StandardScaler** for better model performance.
-

Machine Learning Model

- **Model Used: Random Forest Classifier** (due to its robustness and ability to handle categorical and numerical data).
 - **Training & Testing:** Split data into **80% training** and **20% testing**.
 - **Evaluation Metrics:**
 - **Accuracy:** Assessed overall model performance.
 - **Classification Report:** Checked Precision, Recall, and F1-score.
 - **Confusion Matrix:** Visualized false positives and false negatives.
 - **ROC Curve & AUC Score:** Evaluated model discrimination power.
 - **Feature Importance Analysis:** Identified key factors contributing to churn.
-

Results & Insights

- The **Random Forest model achieved high accuracy** in predicting churn.
 - **Feature importance analysis revealed** that tenure, monthly charges, and contract type were the most significant factors.
 - The **ROC curve showed a good separation between churned and non-churned customers**.
-

Conclusion & Business Impact

- The model successfully predicts customer churn, allowing businesses to **identify high-risk customers**.
- Insights from **EDA and feature importance** can help **optimize customer retention strategies**.

- Businesses can use this model to offer **personalized offers, discounts, or improved services** to reduce churn.