Customer Churn Prediction Using Machine Learning

Author: Garvit Kumbhat

Objective

To predict whether a customer will churn (leave the telecom service) based on their usage, contract, and demographic features using classical machine learning algorithms.

Dataset

- Source: Telco Customer Churn Dataset (Kaggle)
- Rows: ~7,000
- Target: Churn (Yes = 1, No = 0)

Features

- tenure, MonthlyCharges, TotalCharges
- Contract, PaymentMethod, InternetService
- OnlineSecurity, TechSupport
- SeniorCitizen, Partner, Dependents

Data Preprocessing

- Removed customerID
- Converted TotalCharges to numeric
- Handled missing values with median
- Binary-encoded Churn
- One-hot encoded categoricals

Exploratory Data Analysis

- Churn distribution: 73% No, 27% Yes
- Short tenure and high charges = high churn
- Month-to-month contracts = more churn

- Fiber optic users churn more without security

Modeling

Models used:

- Logistic Regression

- Random Forest

- XGBoost

Evaluation:

| Model | Accuracy | Precision | Recall | F1 |

|-------|----------|-----------|--------|----|

| Logistic Regression | 80.2% | 73.1% | 79.5% | 76.1%

| Random Forest | 85.4% | 78.8% | 83.3% | 80.9%

| XGBoost | 87.1% | 81.0% | 85.9% | 83.3% |

Model Interpretation

Top Features:

- tenure

- MonthlyCharges

- Contract_Two year

- InternetService_Fiber optic

- OnlineSecurity_No

Cross Validation

- 5-fold Cross-validation

- Accuracy ~86.3%

- AUC ~0.91

Business Insights

- Short tenure + high charges = high churn

- Two-year contracts = less churn

- Fiber optic with no online security = more churn

Deployment

- Built a Streamlit app with user inputs and predictions

- Live on Streamlit Cloud