

Customer Churn Prediction for Telecom Industry

1. Introduction

Customer churn is a critical metric in the telecom industry, indicating the percentage of customers who discontinue using a service within a given time period. Predicting churn in advance allows telecom providers to proactively take retention measures and reduce revenue loss. This project focuses on developing a machine learning-based classification model that predicts whether a customer is likely to churn based on service usage, demographic data, and account history.

2. Abstract

This project implements a churn prediction system using supervised machine learning. We utilize the Telco Customer Churn dataset from Kaggle, which contains information on customer demographics, services subscribed, tenure, billing, and churn status. The goal is to train a model that classifies whether a customer will churn. Additionally, we use explainable AI (SHAP) to understand the influence of different features on the predictions. The final solution is deployed using a Streamlit web application, allowing users to input customer details and instantly receive churn predictions.

3. Tools & Technologies Used

- Python
- Pandas, Scikit-learn, SHAP, Streamlit
- Random Forest Classifier
- LabelEncoder, StandardScaler
- VS Code, Streamlit (local deployment)

4. Steps Involved in Building the Project

Step 1: Data Collection

- Dataset: Telco Customer Churn (Kaggle)
- 7,043 records with 21 features and a binary target `Churn`.

Step 2: Data Preprocessing

- Dropped `customerID`.

Customer Churn Prediction for Telecom Industry

- Converted `TotalCharges` to numeric and handled missing values.
- Encoded categorical columns using LabelEncoder.
- Scaled numerical columns using StandardScaler.

Step 3: Model Training

- Used RandomForestClassifier and evaluated with classification metrics.

Step 4: Explainability

- Applied SHAP to visualize feature impacts on churn.

Step 5: Web App Interface

- Built using Streamlit for live predictions.

5. Conclusion

This project successfully demonstrates how machine learning can assist telecom companies in understanding and predicting customer churn. The model identifies high-risk customers and aids in retention strategies. SHAP improves trust by explaining model decisions. Streamlit provides a user-friendly interface for real-time predictions.