# Customer Churn Prediction Project

## 1. Dataset Overview

The dataset used in this project is the 'Telco Customer Churn' dataset from IBM. It contains 7043 rows and 21 columns describing customer account information and service usage.

Key Features:

- customerID: Unique identifier.
- gender: Male/Female.
- SeniorCitizen: 0 = No, 1 = Yes.
- tenure: Number of months the customer has stayed.
- MonthlyCharges: Current monthly bill.
- TotalCharges: Total amount charged.
- Churn: Target variable (Yes/No) indicating if the customer left.

Other columns include service options (PhoneService, InternetService, OnlineSecurity, etc.), contract type, and payment method. These categorical features were one-hot encoded before modeling.

Data Cleaning:

- 'TotalCharges' was converted from object to numeric.
- Missing values were handled.
- Features were scaled using StandardScaler.

## 2. Machine Learning Model

We used the XGBoost Classifier due to its strong performance with structured data.

Model Workflow:

- Categorical variables encoded using one-hot encoding.
- Numerical features scaled using StandardScaler.
- The dataset was split into training and testing sets (e.g., 80/20).

- Model was trained to predict the 'Churn' column.

Target Variable:

- 'Churn': 1 if customer is likely to leave, 0 otherwise.

Model Evaluation:

- Evaluated using accuracy, precision, recall, and ROC-AUC.
- The model outputs both a class label and a probability of churn.

Important Predictive Features:

- Contract type, MonthlyCharges, Tenure, InternetService, and PaymentMethod showed high impact on churn.

## 3. Streamlit App Interface

We built an interactive app using Streamlit to make churn predictions in real-time.

App Features:

- Collects customer information (tenure, charges, contract type, etc.) via widgets.
- Constructs a feature vector in the format used during training.
- Applies the same scaler and feeds data into the trained XGBoost model.
- Outputs a churn prediction with probability.

The app allows businesses to proactively identify customers likely to churn and take preventive action. It can be deployed locally or on platforms like Streamlit Cloud.