Predicting Customer Churn for a Telecom Company

Submitted by: Usha Rani

# 1. Introduction

Customer churn is a major concern for telecom companies as it directly impacts their revenue and growth. Predicting customer churn helps in identifying customers likely to discontinue services, allowing businesses to take preventive actions. This project involves building a machine learning model to predict customer churn based on demographic, service usage, and account information.

# 2. Objective

To build and deploy a machine learning model capable of predicting whether a customer will churn or not, based on relevant features such as tenure, monthly charges, total charges, gender, contract type, internet service, and payment method.

# 3. Data Description

The dataset contains the following customer attributes:  
- Tenure: Number of months the customer has stayed with the company  
- MonthlyCharges: Monthly fee the customer pays  
- TotalCharges: Total amount charged to the customer  
- Gender: Male or Female  
- InternetService: Type of internet service (DSL, Fiber optic, None)  
- Contract: Type of contract (Month-to-month, One year, Two year)  
- PaymentMethod: Customer's payment method (Electronic check, Mailed check, etc.)  
- Churn: Target variable indicating whether the customer left (Yes) or not (No)

# 4. Methodology

The project workflow includes:  
1. Data preprocessing (handling missing values, encoding categorical variables)  
2. Creating a pipeline that automates preprocessing and model training  
3. Splitting data into training and testing sets  
4. Training models (e.g., Logistic Regression, Random Forest)  
5. Evaluating the best-performing model  
6. Saving the trained pipeline using joblib

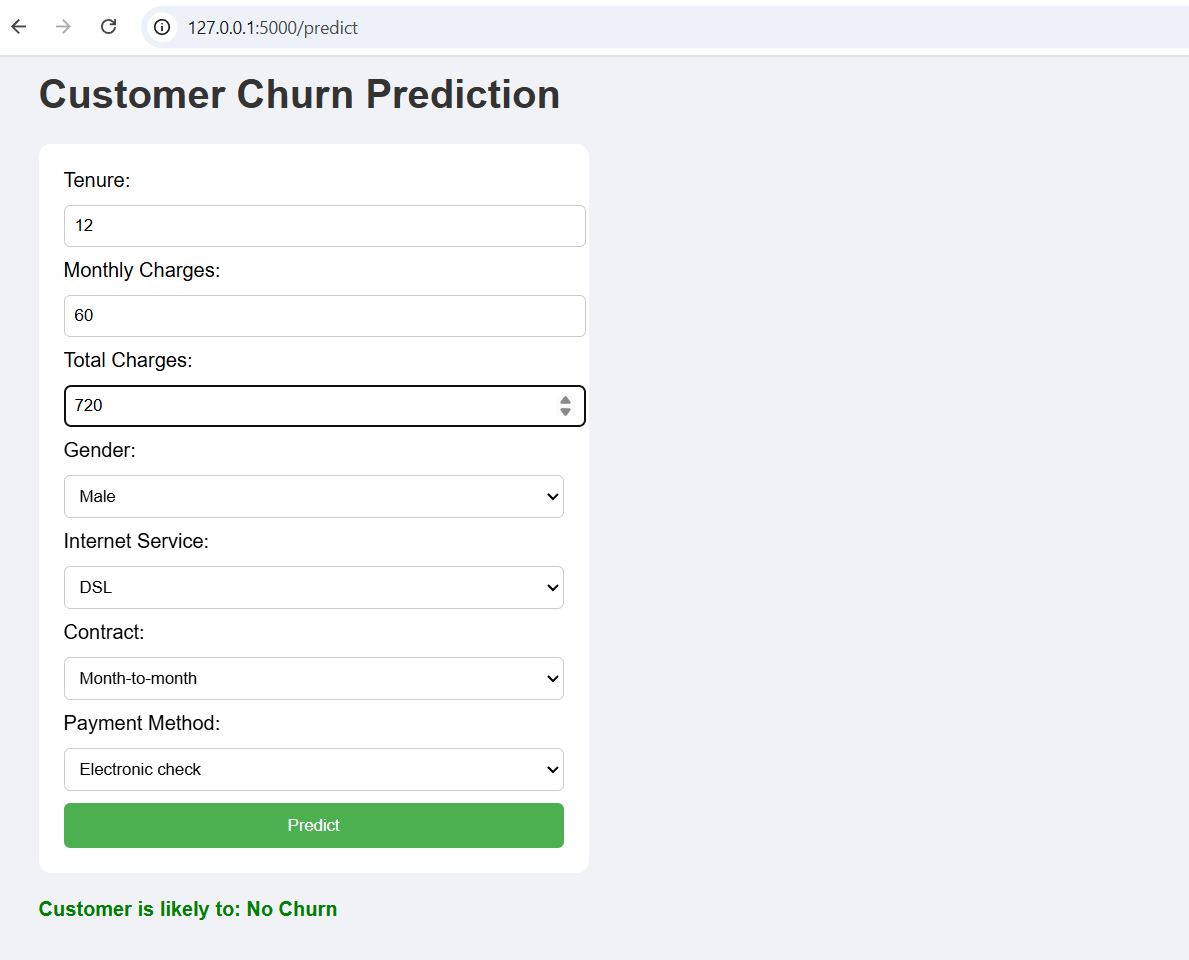
7. Creating a Flask web application to deploy the model

# 5. Code and Implementation

The implementation involved the following components:  
  
- A pipeline combining preprocessing steps with a classifier:  
 • ColumnTransformer for handling numerical and categorical features  
 • OneHotEncoder for categorical data  
 • SimpleImputer for missing values  
 • Logistic Regression or Random Forest as classifier  
  
- The trained pipeline was saved using joblib:  
 model = Pipeline([...])  
 joblib.dump(model, 'churn\_model\_pipeline.pkl')  
  
- The Flask app (`app.py`) accepts form inputs and uses the pipeline to make predictions  
- An HTML form (`index.html`) collects user input for prediction

# 6. Results and Observations

The model was tested using various combinations of input. The Flask web application predicts whether a customer is likely to churn. Below is a screenshot of the working application:



# 7. Conclusion

This project successfully demonstrates the end-to-end process of building and deploying a machine learning model to predict customer churn. The web app enables real-time predictions using a user-friendly interface, empowering telecom companies to proactively engage with at-risk customers.