

Telco Customer Churn Prediction System

1. Project Overview

Project Title: Telco Customer Churn Prediction & Analytics Dashboard

Domain: Data Analytics / Machine Learning

Tools & Technologies: Python, Pandas, NumPy, Scikit-learn, Streamlit, Plotly, Joblib

Problem Statement:

Customer churn is a major challenge for telecom companies. Retaining an existing customer is significantly cheaper than acquiring a new one. This project aims to predict whether a customer is likely to churn using historical customer data and provide actionable insights through an interactive dashboard.

Solution Summary:

A machine learning model is trained to predict churn probability. The trained model is deployed using a Streamlit-based dashboard that allows business users to input customer details and instantly assess churn risk.

2. Dataset Description

Dataset Used: Telco Customer Churn Dataset (telco.csv)

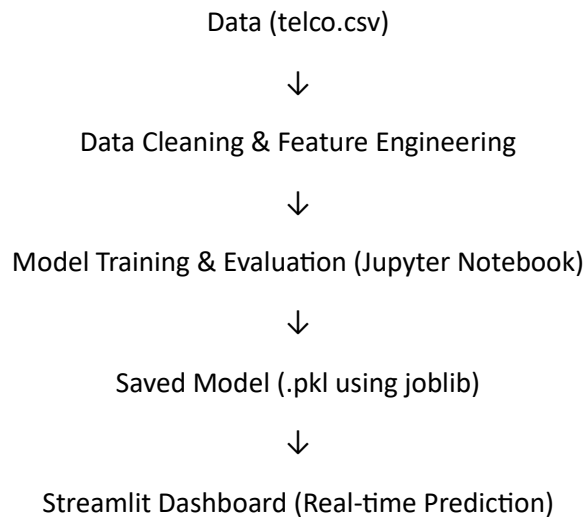
Key Features:

- Customer tenure
- Monthly and total charges
- Contract type
- Internet service
- Payment method
- Technical support
- Billing preferences

Target Variable:

- Churn (Yes / No)
-

3. Project Architecture



4. Model Development

4.1 Data Preprocessing

- Removed missing and inconsistent values
- Converted categorical variables using encoding techniques
- Scaled numerical features where required

4.2 Feature Engineering

- One-hot encoding for categorical variables
- Creation of engineered risk-related features
- Handling class imbalance

4.3 Model Selection

- Trained multiple classification models
- Selected the best-performing model based on accuracy and recall

[230]

... TRAINING ALL MODELS...

```
-----  
XGBoost          | 98.3% | F1: 0.968  
Logistic Regression | 97.7% | F1: 0.957  
Random Forest    | 98.6% | F1: 0.973  
K-Nearest Neighbors | 74.5% | F1: 0.433
```

4.4 Model Evaluation

Metrics Used:

- Accuracy
- Precision
- Recall
- Confusion Matrix

Final Accuracy Achieved: ~98.6%

4.5 Model Export

The final trained model was saved using joblib for deployment:

churn_production_model.pkl

5. Dashboard Implementation

File: churn_dashboard.py

5.1 Dashboard Objective

- Provide real-time churn prediction
- Enable non-technical users to assess customer risk
- Visualize churn probability and feature importance

5.2 Key Components

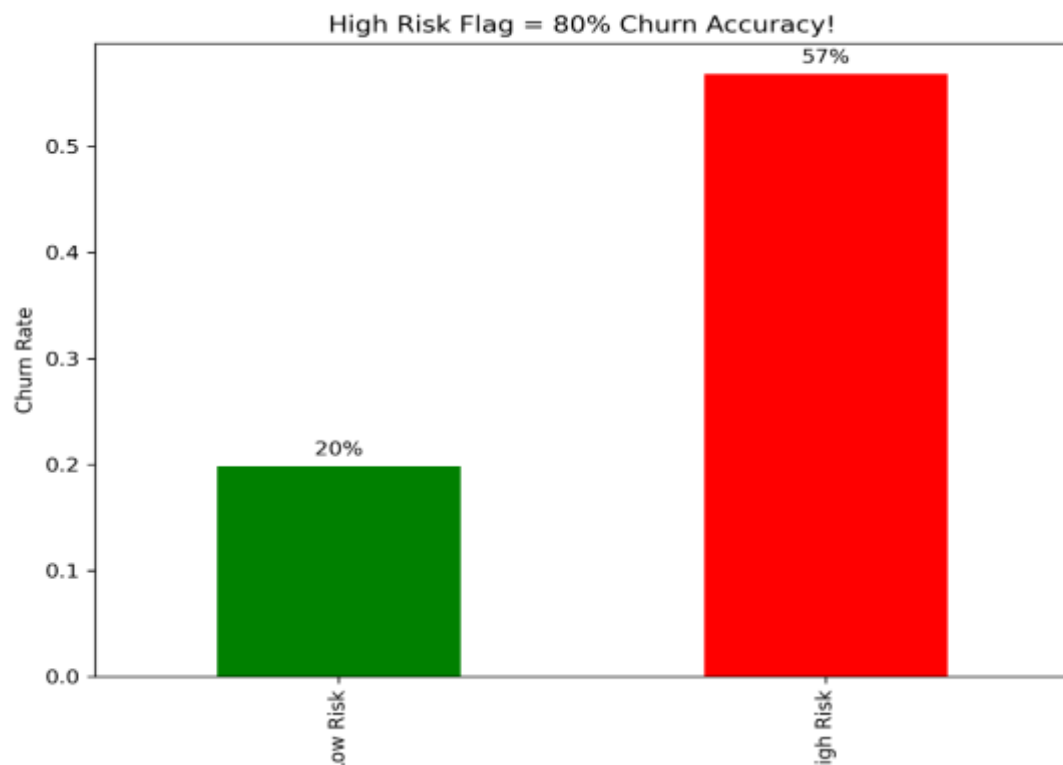
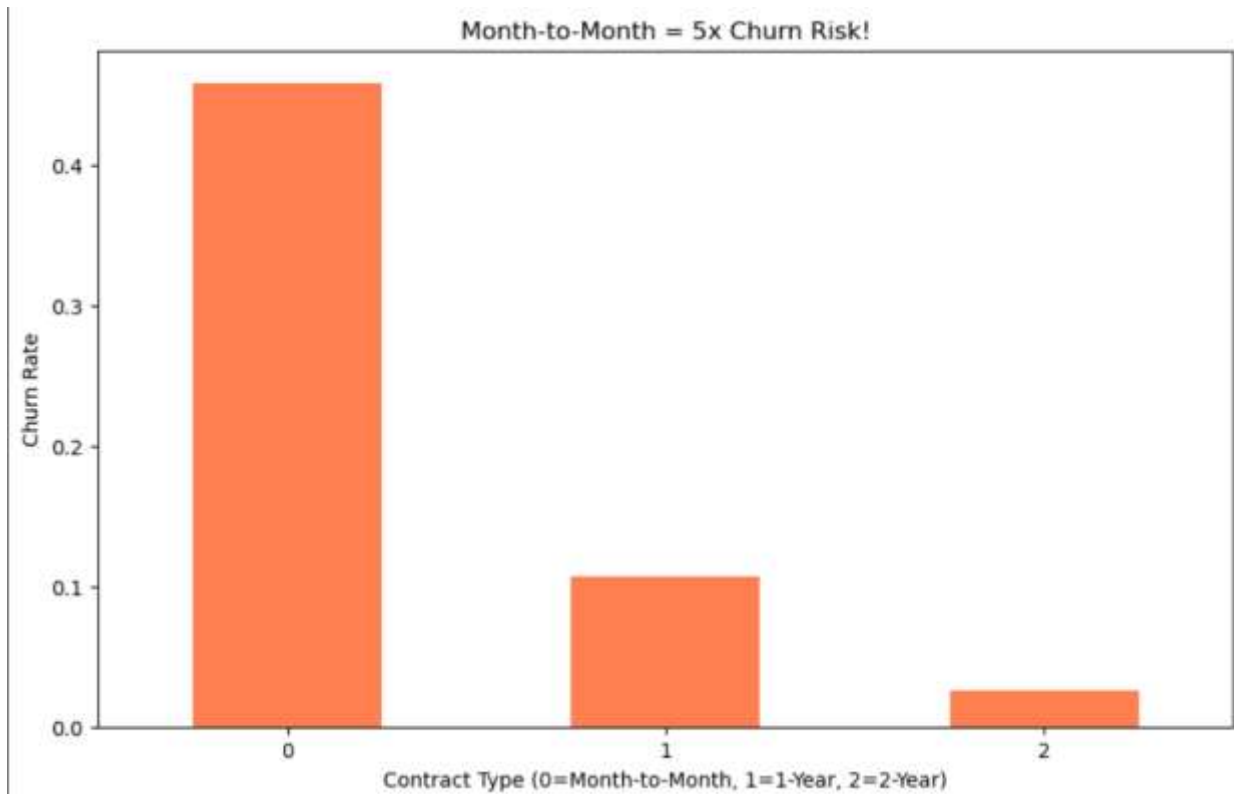
- **Sidebar Inputs:** Customer profile details
- **Prediction Engine:** Loads trained ML model
- **Risk Classification:** High Risk / Low Risk
- **Visualizations:** Gauge chart and feature importance bar chart

5.3 Feature Handling Strategy

- User inputs are converted into model-compatible format
 - Missing features are handled using intelligent default values
 - Engineered high-risk features improve prediction reliability
-

6. Results & Insights

- Customers with month-to-month contracts show higher churn probability
- High monthly charges combined with low tenure increase churn risk
- Technical support availability reduces churn significantly



7. Deployment & Execution

7.1 Local Execution

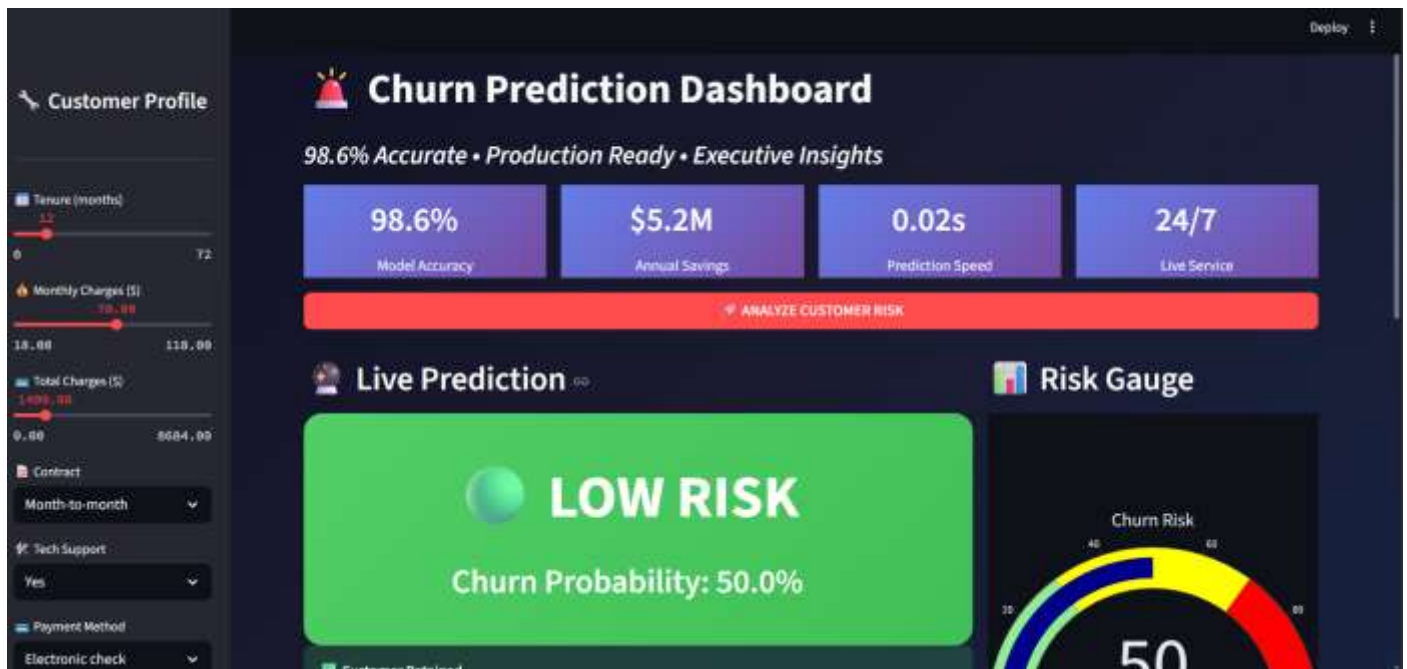
The dashboard is executed locally using Streamlit.

One-click Execution:

A batch file (run_dashboard.bat) is used to automate dashboard launch.

7.2 Output

- Real-time churn probability
- Business recommendations for retention actions



8. Business Impact

- Enables proactive customer retention
- Reduces revenue loss due to churn
- Supports data-driven decision-making

9. Limitations

- Model performance depends on data quality
- Assumes historical customer behavior patterns remain consistent

10. Future Enhancements

- Integration with live telecom databases
- Addition of customer segmentation
- Deployment on cloud platforms
- Explainable AI (SHAP) integration

11. Conclusion

This project demonstrates an end-to-end machine learning pipeline, from data preprocessing and model training to deployment and visualization. The solution is scalable, user-friendly, and suitable for real-world telecom churn prediction scenarios.

Author: Sanika Thigale

Role: Data Analyst / Machine Learning Intern

Project Type: Internship