Project Files Overview

1. Telecom-Customer-Churn-Prediction.ipynb

Purpose: The main Jupyter Notebook containing the complete data science workflow — from data loading, cleaning, visualization, and feature analysis to training multiple machine learning models.

Models Used: Logistic Regression, Random Forest, SVM, AdaBoost, and XGBoost.

Outcome: SVM was identified as the best-performing model and selected for deployment.

Key Insight: Contract type, tenure, and monthly charges are the strongest predictors of customer churn.

2. churn_main_pr.py

Purpose: Python training script that reproduces the final model training pipeline.

Functionality:

- Loads the dataset (`Telco_Customer_Churn.csv`)
- Encodes categorical variables (`InternetService`, `Contract`, `Churn`)
- Selects key features: `tenure`, `InternetService`, `Contract`, `MonthlyCharges`, and `TotalCharges`
- Trains a Support Vector Machine (SVM) classifier with a linear kernel
- Saves the model as `svm_model.joblib` for deployment

Insight: Ensures consistent preprocessing and model generation independent of the notebook.

3. svm_model.joblib

Purpose: The saved, trained SVM model used for prediction.

Generated by: `churn_main_pr.py`

Usage: Loaded by the Streamlit app to make real-time churn predictions.

Insight: Captures the best churn classification performance from the experiments.

4. churn_app_pr.py

Purpose: Streamlit-based deployment app for interactive churn prediction.

Features:

- Accepts customer data inputs: tenure, contract, internet service, monthly and total charges
- Automatically encodes categorical values
- Loads the trained model (`svm_model.joblib`)
- Displays prediction results:
  - green**This customer is likely to stay.**
  - red**This customer is likely to churn.**

Insight: Converts the trained ML model into a user-friendly predictive web application.