TELECOM OPERATOR CUSTOMER CHURN PREDICTOR

Project 3

CONTENTS

PROJECT DESCRIPTION

A data-driven model created and trained on Python to identify customer churn for a telecom service provider, enabling companies to take proactive retention actions.

It enables telecom operators to act early and effectively, minimizing customer loss and optimizing retention strategies.

PYTHON SCRIPT PIPELINE

LOADING DATASET

- The script begins by loading the customer data from a file named 'telecom_churn.csv'.

CLEANING DATA

- The non-predictive *customerID* column is removed.
- The *TotalCharges* column is converted to a numerical format, and any missing values are filled with the median charge to avoid errors.

ENCODING DATA

- All categorical text data (like 'gender', 'Yes', 'No') is converted into numerical values using *LabelEncoder*. This is necessary because machine learning models require numerical input.

PYTHON SCRIPT PIPELINE

TRAINING SET (80%)

- Used to teach the model the patterns associated with customer churn.

TESTING DATA (20%)

- Kept aside to evaluate the model's performance on unseen data.

TRAIN THE MODEL

- A **Random Forest Classifier**, which is an ensemble of decision trees, is trained on the training data. The model learns the relationships between customer attributes and the likelihood of churning.

PYTHON SCRIPT PIPELINE

MAKE PREDICTIONS

- The trained model is used to make predictions on the testing set.

ASSESS ACCURACY

- The model's predictions are compared to the actual outcomes in the test data to calculate its **accuracy**.

GENERATE REPORTS

- A **Classification Report** is created to show more detailed metrics like precision and recall for both "Churn" and "No Churn" predictions.
- A **Confusion Matrix** is generated to visualize how many predictions were correct and incorrect.

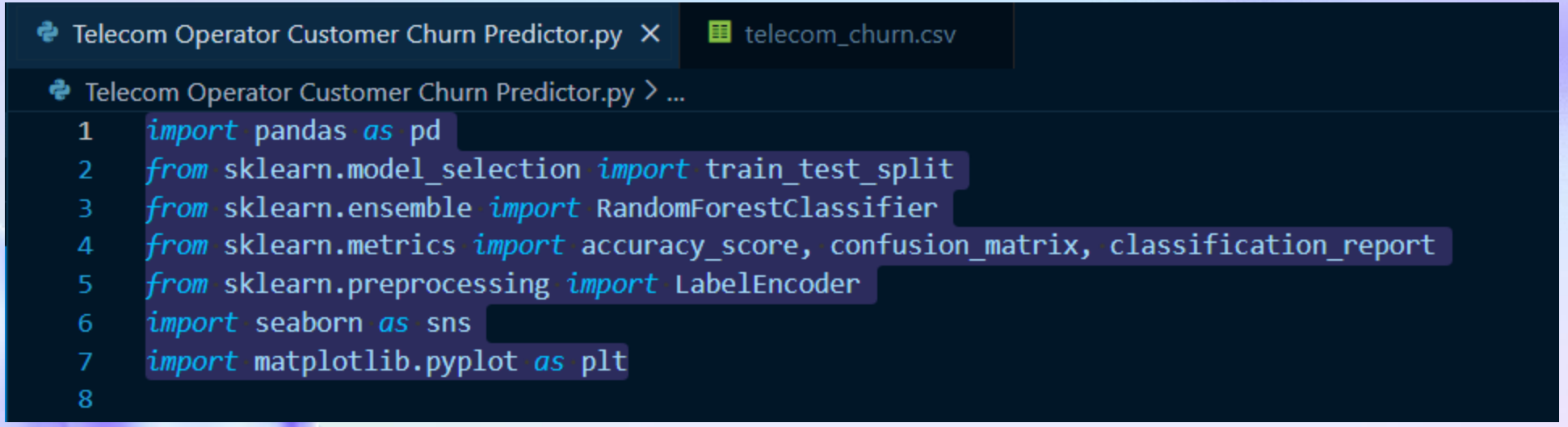
PYTHON SCRIPT PIPELINE

RESULTS & INSIGHTS

Visualize Key Factors:

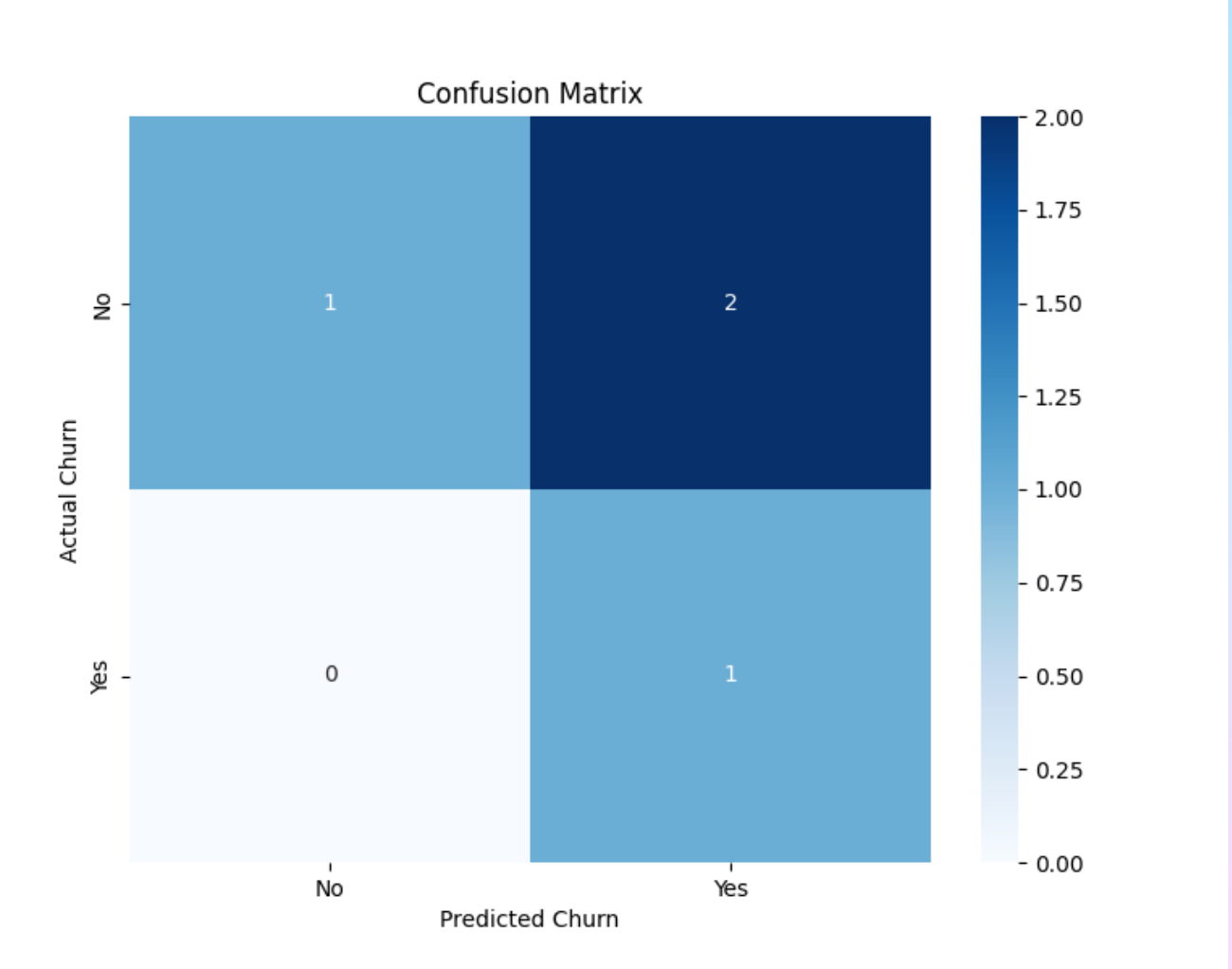
The script identifies and plots the **Top 10 most important features** (e.g., contract type, tenure, monthly charges) that most influence a customer's decision to churn. This helps in understanding the key drivers behind customer churn.

USE OF LIBRARIES



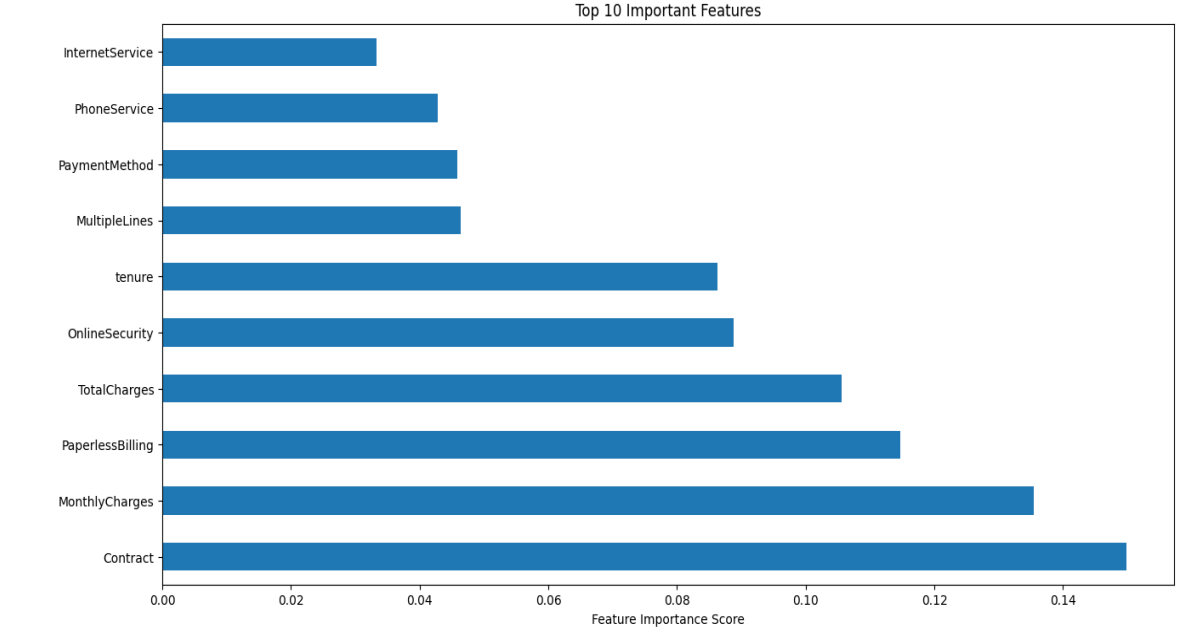
USE OF HEATMAP

The heatmap tells us the confusion matrix gives a detailed breakdown of the model's performance by showing where it was right and where it was wrong. It's a grid comparing the **Actual Churn** (the truth) against the **Predicted Churn** (the model's guess).



USE OF BAR PLOT

The bar plot ranks the most influential factors that drive customer churn, according to your model. The longer the bar, the more important the feature. This model conclude that they should focus their efforts on customers with specific **contract types** and **high monthly charges**.



I hope the presentation met your expectations, and sorry for any inconvenience if caused.

THANK YOU

Nisarg Kumar Gharde