****

**Assessment Report**

**on**

**“Classify Customer Churn”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

in

**Intro to AI**

By

Tushar Bhardwaj (202401100300265,CSE(AI)-D)

**Under the supervision of**

Abhishek Shukla.

**KIET Group of Institutions, Ghaziabad**

**May, 2025**

**1. Introduction**

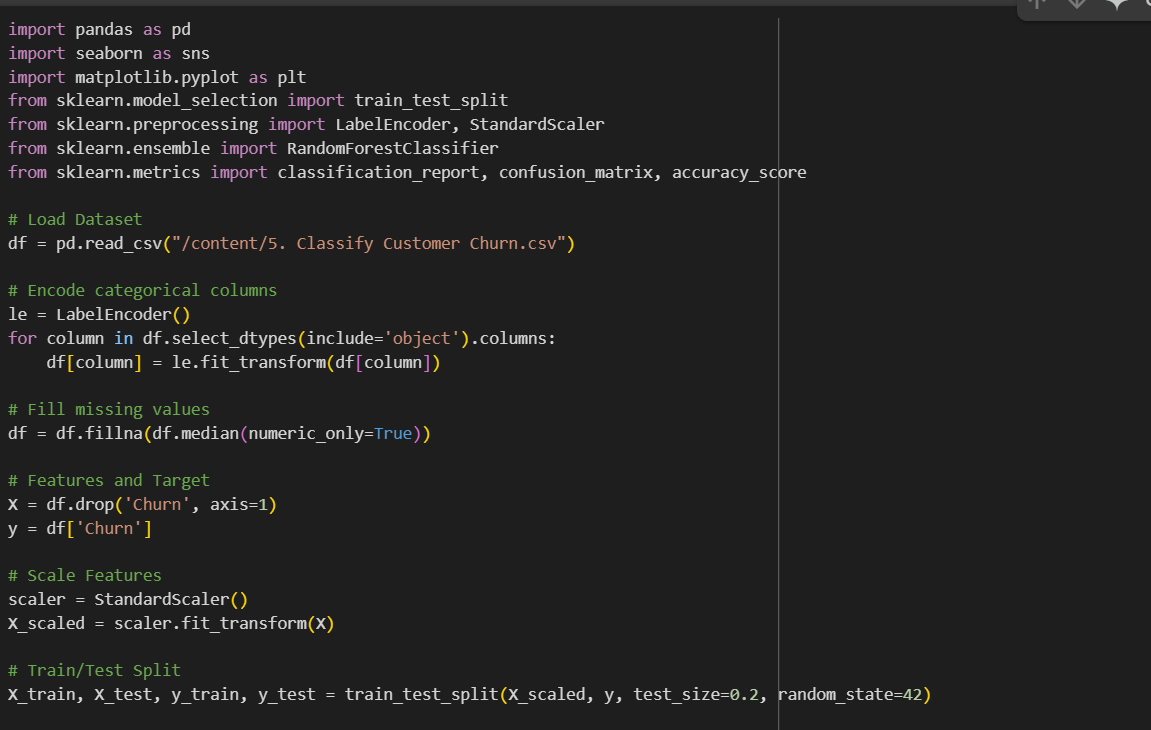
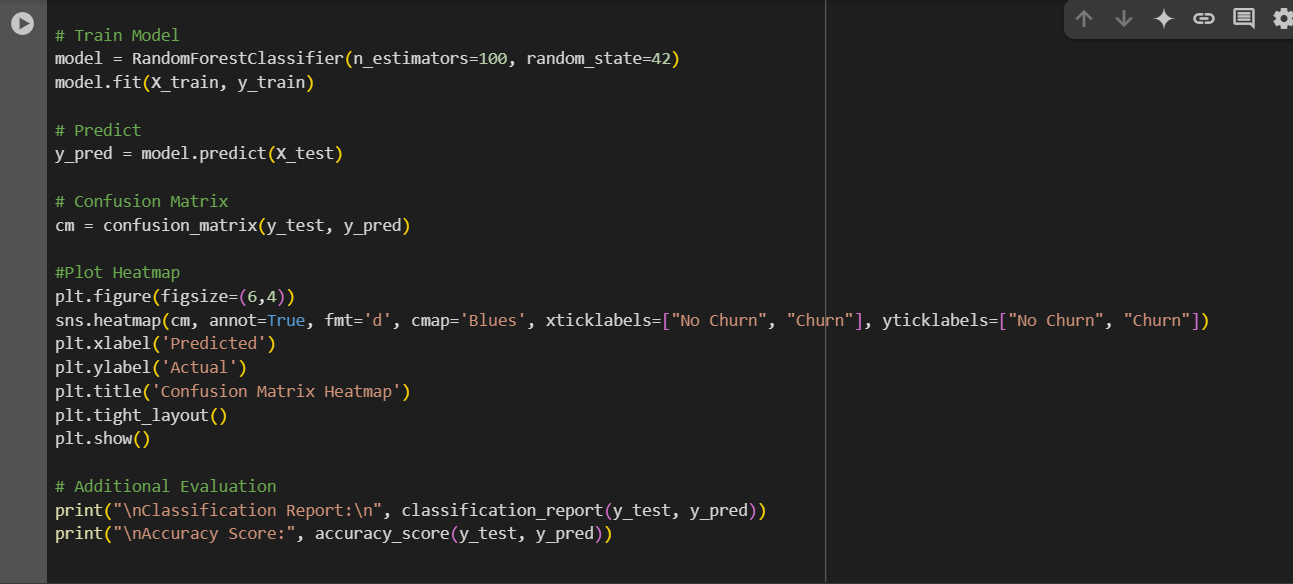
Customer churn prediction is a vital task for businesses that want to retain their customers and maintain revenue streams. In this project, we use machine learning techniques to classify whether a customer is likely to churn or not based on historical data. This classification task helps businesses to identify potential churners and take preventive measures to improve customer retention.

**2. Dataset Description**

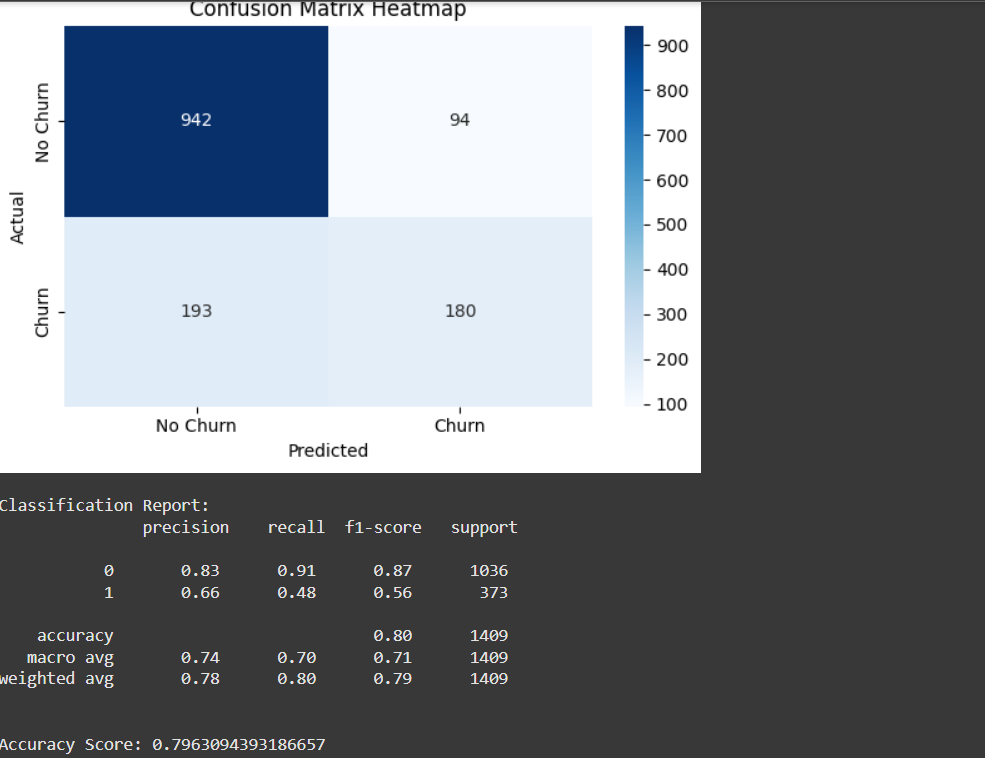
The dataset contains customer-related information such as demographic data, account details, and service usage. It includes both categorical and numerical variables. The target column `Churn` is binary and indicates whether the customer has churned (`1`) or not (`0`).  
  
Key features in the dataset include:  
- Gender  
- SeniorCitizen  
- Partner  
- Dependents  
- Tenure  
- PhoneService  
- MultipleLines  
- InternetService  
- Contract  
- MonthlyCharges  
- TotalCharges  
... and many others.

**3. Methodology**

The following steps were performed to build and evaluate the machine learning model:  
  
1. Data Preprocessing:  
 - All categorical columns were encoded using LabelEncoder.  
 - Missing values were filled using median imputation for numeric columns.  
  
2. Feature Scaling:  
 - Features were scaled using StandardScaler to normalize their values.  
  
3. Train-Test Split:  
 - The dataset was split into training and test sets in an 80:20 ratio.  
  
4. Model Training:  
 - A RandomForestClassifier was trained using 100 estimators (trees).

  
  
5. Evaluation:  
 - The model was evaluated using a Confusion Matrix and metrics like accuracy, precision, recall, and F1-score.

**4. Evaluation Metrics**



**5. Conclusion**

The Random Forest classifier achieved an accuracy of approximately 79.6%. While it performs well in predicting non-churned customers, further improvements can be made in predicting churned ones by tuning hyperparameters or using techniques like SMOTE for class balancing. This model can be a strong asset for companies looking to proactively address customer churn.

**6. References**

• scikit-learn documentation  
• matplotlib and seaborn for visualization  
• Dataset source: Provided CSV  
• Open-source examples of classification models

**7. Developer Details**

Name: Tushar Bhardwaj