

HSBC HACKATHON

Problem Statement: 1

Description

Develop an AI model capable of detecting fraudulent transactions in real-time. Use historical transaction data to train the model to identify anomalies and flag.

Objective

Successfully implement and deploy a model that can accurately detect fraudulent transactions with minimal false positives.

Strategy

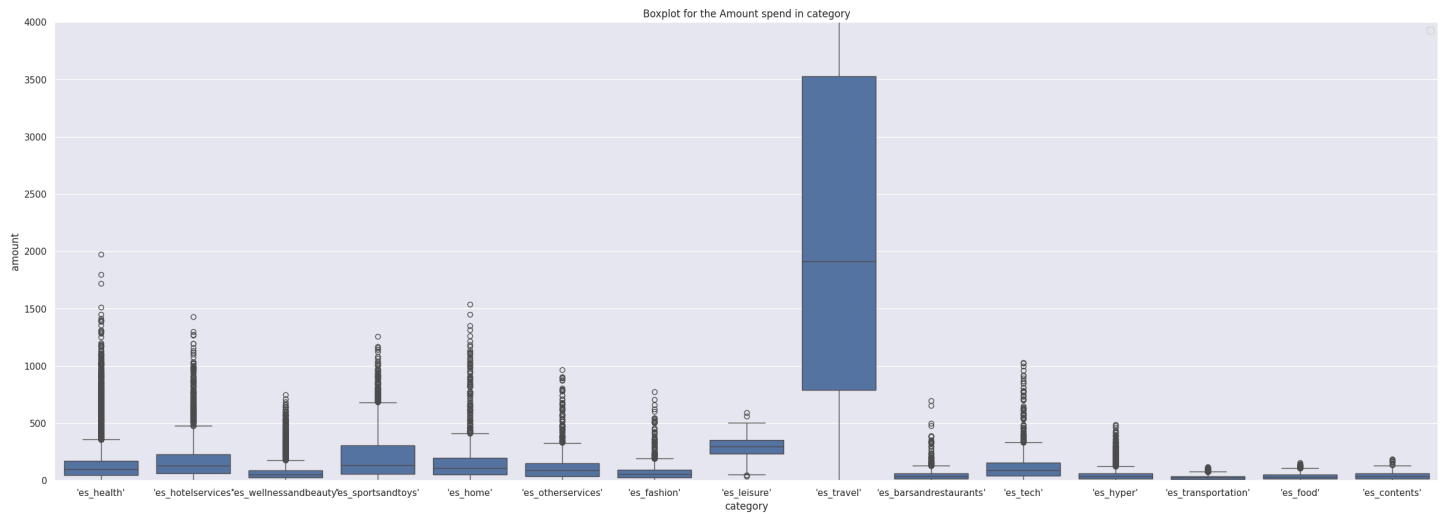
- Exploratory Data Analysis
- Data Preprocessing
- Model Fitting

Exploratory Data Analysis

- 1) Amount is the most correlated feature to fraud
- 2) Dataset is extremely skewed

```
fraud
0    507443
1      6200
Name: count, dtype: int64
```

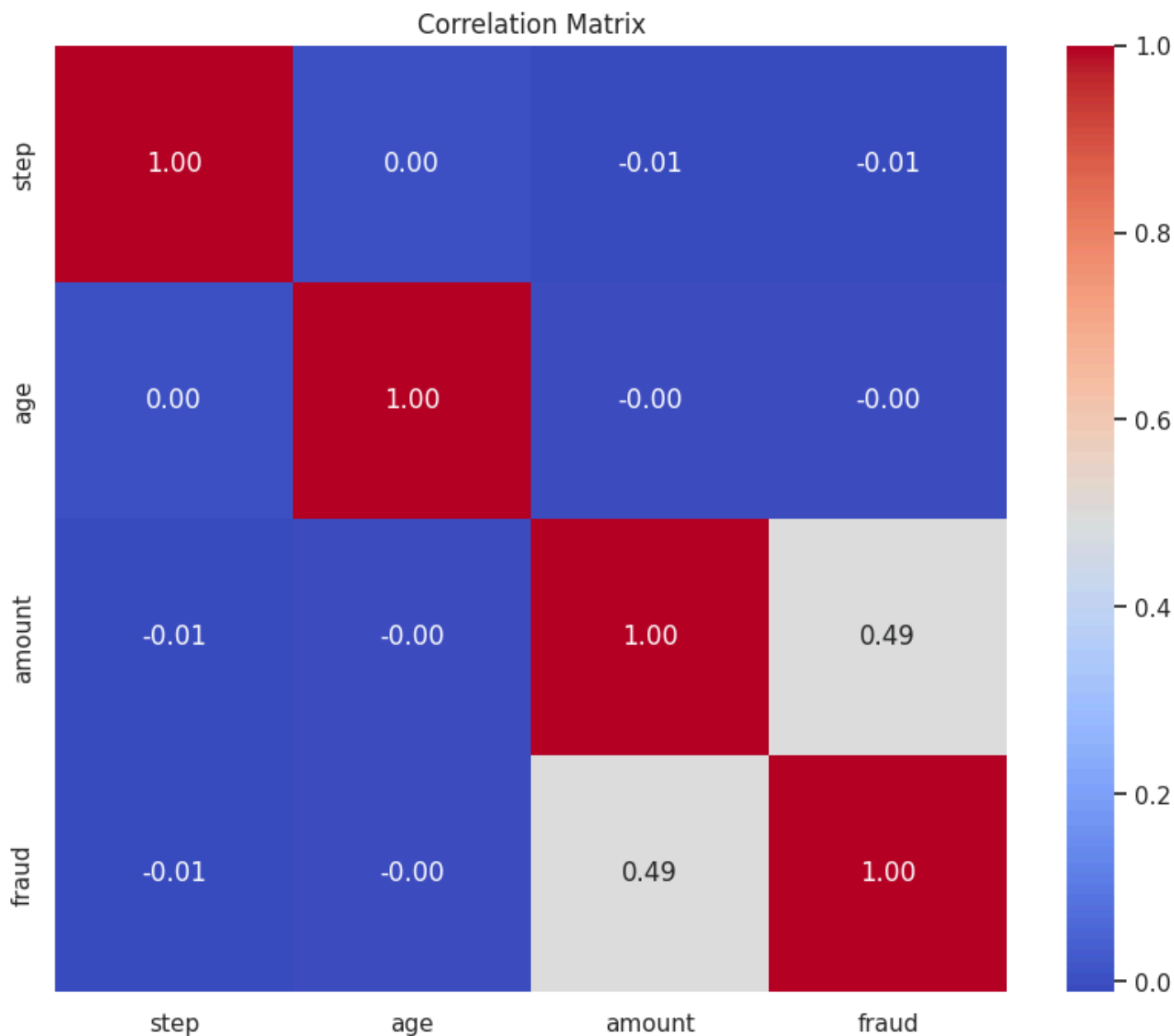
3) Category wise analysis: Leisure and travel observe highest number of frauds (Customers spend more money on these).



4) Age 'O' has highest number of frauds

5) Gender 'F' has highest number of frauds

6) Zipcode is constant everywhere so columns are dropped



Model Details (Ensemble Learning)

Random Forest

```

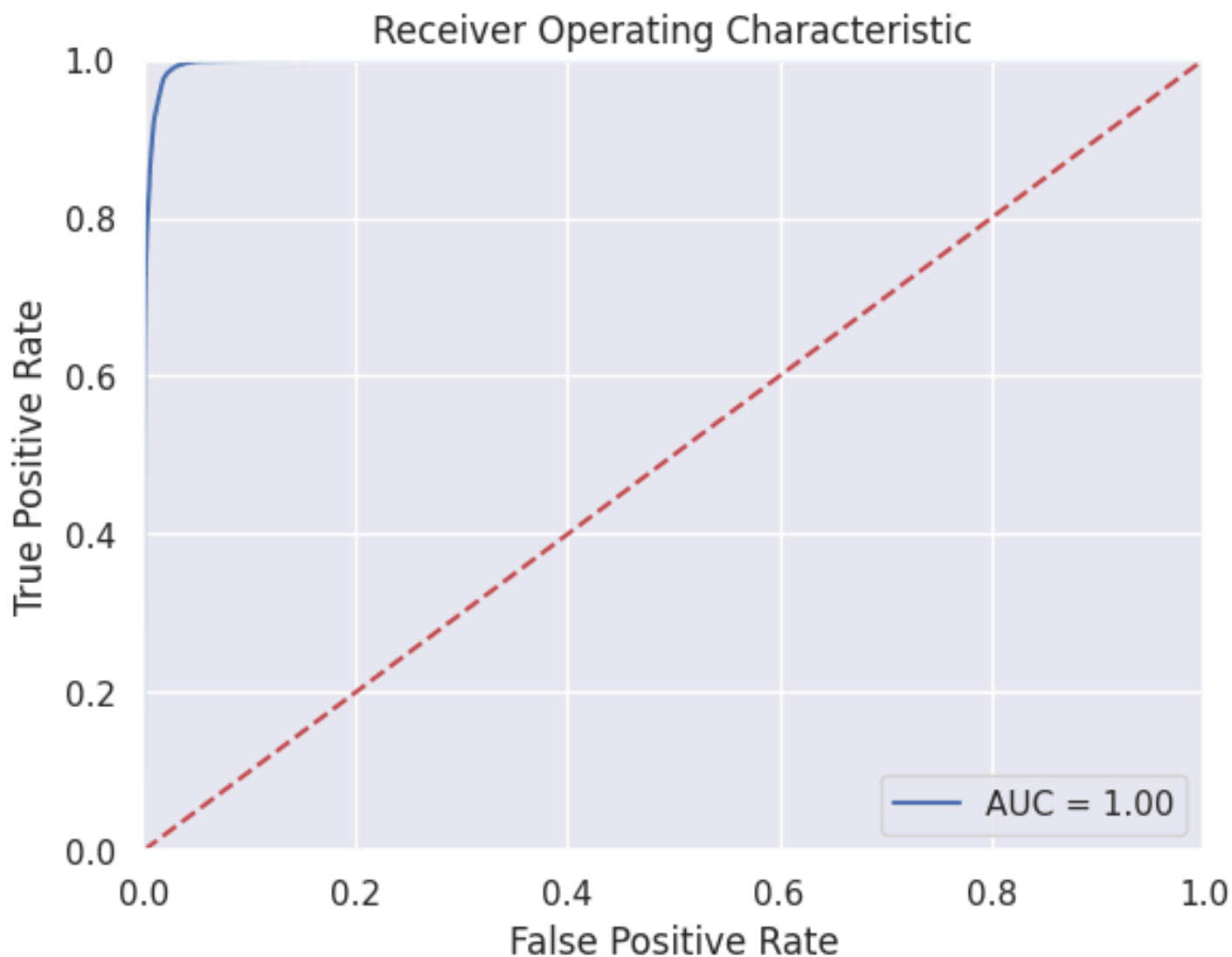
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:1152: DataConversionWarning: A column-vector y was passed when a
return fit_method(estimator, *args, **kwargs)
[Parallel(n_jobs=1)]: Done 49 tasks      | elapsed: 45.5s
[Parallel(n_jobs=1)]: Done 49 tasks      | elapsed: 1.6s
Classification Report for Random Forest Classifier:
      precision    recall  f1-score   support

     0       0.99      0.97      0.98     152233
     1       0.97      0.99      0.98     152233

 accuracy          0.98          0.98          0.98     304466
 macro avg          0.98          0.98          0.98     304466
weighted avg          0.98          0.98          0.98     304466

Confusion Matrix of Random Forest Classifier:
[[146936  5297]
 [   911 151322]]
[Parallel(n_jobs=1)]: Done 49 tasks      | elapsed: 1.2s

```



XGBoost

```
Classification Report for XGBoost:
      precision    recall  f1-score   support

     0       1.00      0.99      0.99     152233
     1       0.99      1.00      0.99     152233

 accuracy      0.99
 macro avg     0.99      0.99      0.99     304466
 weighted avg   0.99      0.99      0.99     304466

Confusion Matrix of XGBoost:
[[150483  1750]
 [   709 151524]]
```

Receiver Operating Characteristic

