BVRIT HYDERABAD College Of Enigneering for Women

CREDIT CARD FRAUD DETECTION

DEPARTMENT OF CSE(AI & ML)

Team No: 11

20WH1A6654 - C Sucharitha

20WH1A6655 - N Manvitha

20WH1A6656 - M Sonali Varma

20WH1A6657 - Y Manaswini

20WH1A6658 - A Vyshnavi

April 1, 2023

PROBLEM STATEMENT

- It is important that credit card companies fraudulent credit card transactions so that customers are not charged for items that they did not purchase.
- Hence We Predict Whether Provided Transaction is Fraudulent or Valid.

PYTHON PACKAGES AND LIBRARIES

- Numpy
- Pandas
- Matplotlib
- Seaborn
- Scikit-learn
- XGBoost
- TKinter

Types of Algorithms

- **Naive Baye's Classifier**
- Support Vector Machine
- Random Forest Classifier
- Linear Regression
- K Neighbours Classifier
- XGB Classifier

Explanation

- We have chosen Classification algorithm.After plotting ROC for above mentioned classifiers,we obtained maximum ROC-AUC for Navie Baye's Classifier.
- Therefore out of 6 classifiers tested,we have picked Navie Baye's Classifier.

Naive Baye's Classifier

- Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.
- It is mainly used in text classification that includes a high-dimensional training dataset.
- Accuracy: 0.978810435
- ROC-AUC: 0.92035306847

Support Vector Classifier

- Support vector Machine algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future.
- SVM algorithm can be used for Face detection, image classification, text categorization.
- Accuracy:0.999350444
- ROC-AUC:0.909398622266

Random Forest Classifier

- Random Forest is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.
- The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.
- Accuracy: 0.999578666
- ROC-AUC:0.90951292204

Logistic Regression Classifier

- Logistic regression is used for predicting the categorical dependent variable using a given set of independent variables.
- Logistic Regression is a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets.
- Accuracy: 0.999280221
- ROC-AUC:0.84032881671

K Neighbours Classifier

- K-Nearest Neighbour algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data.
- Accuracy: 0.999455777
- ROC-AUC:0.893520306076

XGB Classifier

- XGBoost is an ensemble learning method that combines the predictions of multiple weak models to produce a stronger prediction.
- XGBoost is efficient in handling of missing values, which allows it to handle real-world data with missing values without requiring significant pre-processing.
- Accuracy: 0.999666444
- ROC-AUC:0.920177596793

Comparison Table

S.NO	CLASSIFIER	ACCURACY	PRECISION	RECALL	F1-SCORE	ROC-AUC
1	Naïve Bayes	0.978810435	0.06352941176	0.86170212765	0.11833455	0.92035306847
2	Random Forest	0.999578666	0.91666666666	0.81914893617	0.86516853	0.90951292204
3	Logistic Regression	0.999280221	0.85333333333	0.68085106382	0.75739644	0.84032881671
4	K Neighbours	0.999455777	0.87058823529	0.78723404255	0.82681564	0.893520306076
5	XGB	0.999666444	0.95180722891	0.84042553191	0.89265536	0.920177596793
6	Support Vector Machine	0.999350444	0.79381443298	0.81914893617	0.80628272	0.909398622266

Figure: Models Comparison

ROC-AUC

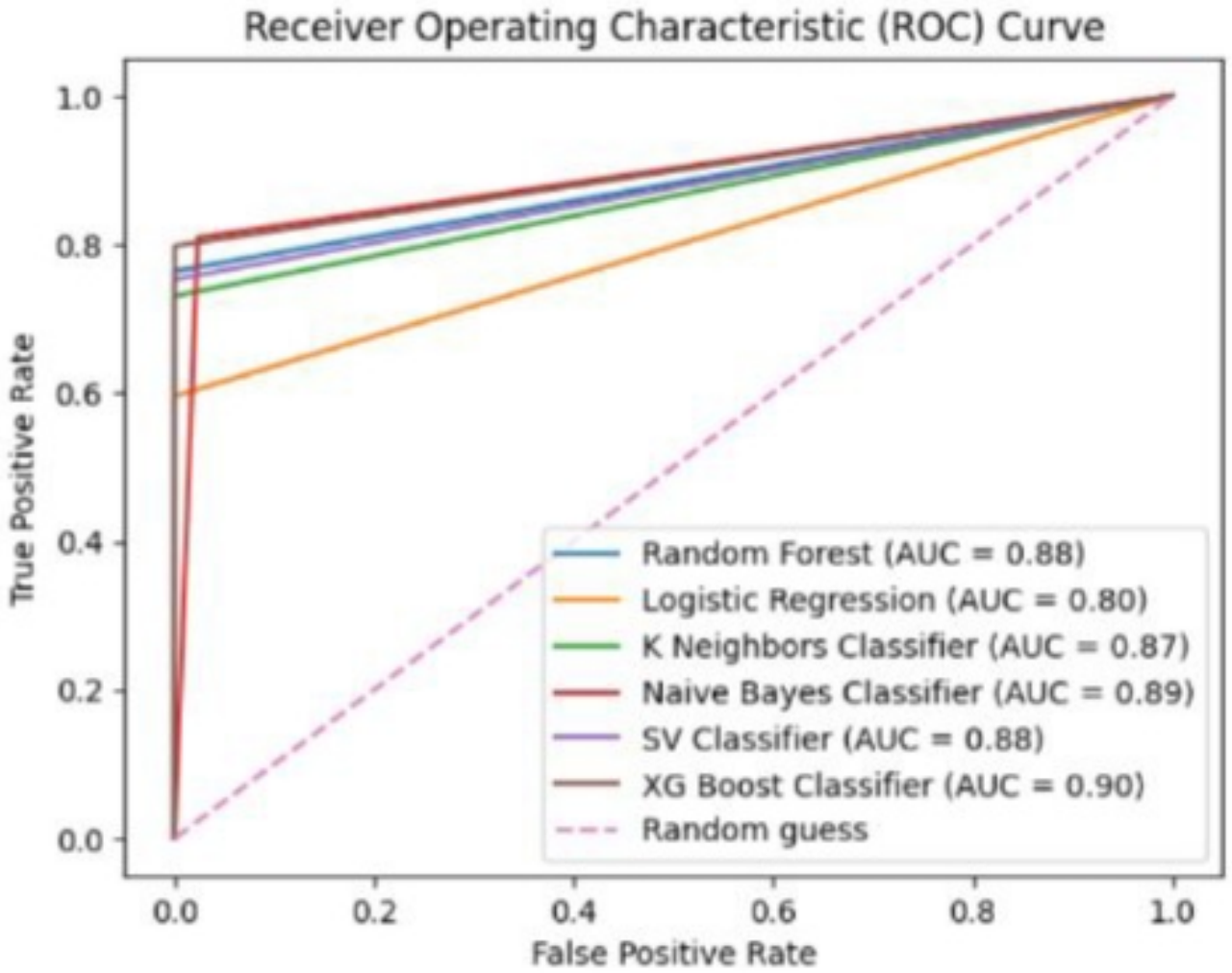
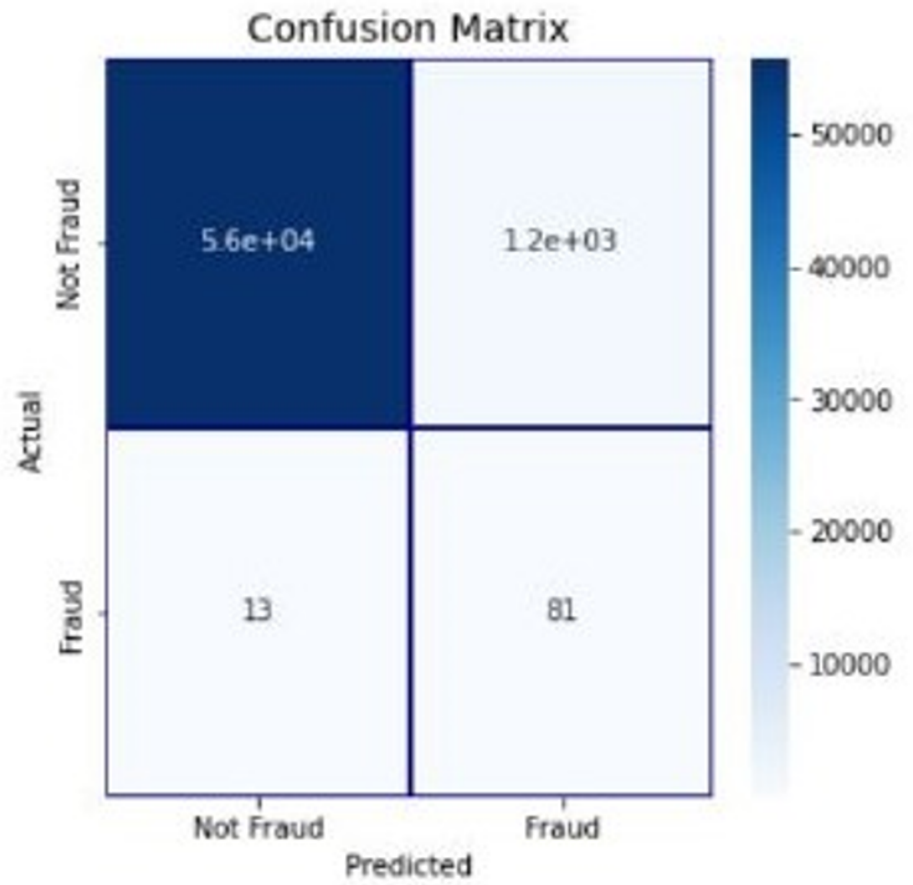


Figure: Caption

Confusion Matrix



OUTPUT

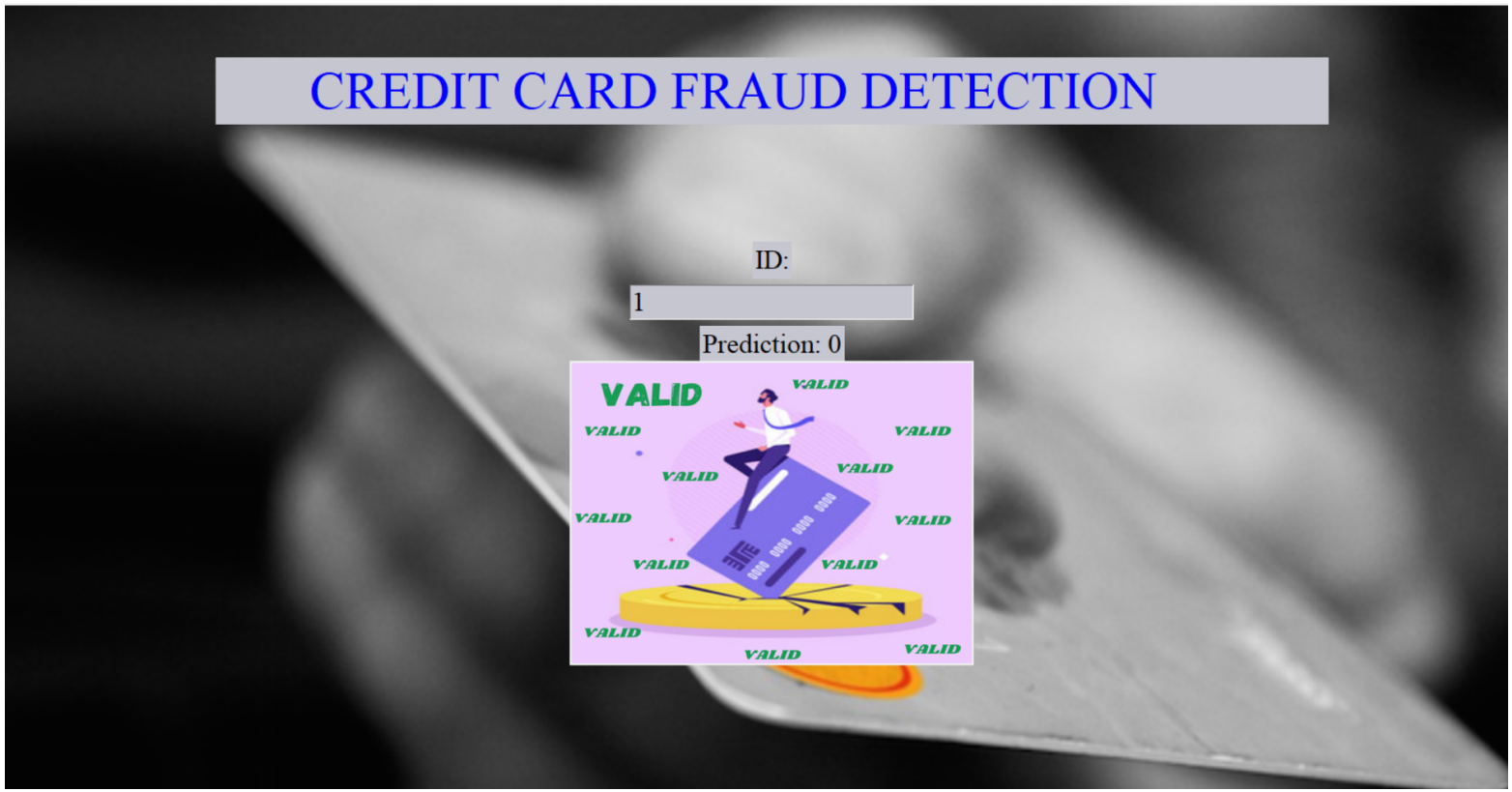


Figure: Output

THANK YOU