

Digital Payment Fraud Detection

Overview:

Digital payment fraud is a significant concern in today's digital world. Digital payment systems have helped a lot in the ease of payments. But, at the same time, it increased in payment frauds. Digital payment frauds can happen with anyone using any payment system, especially while making payments using a credit card. That is why detecting Digital payment fraud is very important.

Objective:

This project aims to develop a fraud detection system using machine learning algorithms.

Data Collection:

Dataset Columns:

- step: represents a unit of time where 1 step equals 1 hour.
- type: type of online transaction.
- amount: the amount of the transaction.
- nameOrig: customer starting the transaction.
- oldbalanceOrg: balance before the transaction.
- newbalanceOrig: balance after the transaction.
- nameDest: recipient of the transaction.
- oldbalanceDest: initial balance of recipient before the transaction.
- newbalanceDest: the new balance of recipient after the transaction.
- isFraud: fraud transaction.

Data Preprocessing:

1. **Data Cleaning:** Identify Null Values: Begin by identifying columns or features with Null values.
2. **Handling Outliers:** Check any outliers present or not using box plot.
3. **Standardization:** Scale numerical features to a standard range (e.g., 0-1)
4. **Encoding:** Apply one hot encoding to Convert categorical variables into a binary format, creating separate columns for each category.
5. **Feature Engineering:** Create new features that might be more predictive or meaningful for the model (e.g. nameOrig, nameDest).

Exploratory Data Analysis (EDA): Exploratory Data Analysis or EDA is used to take insights from the data. Data Scientists and Analysts try to find different patterns, relations, and anomalies in the data using some statistical graphs and other visualization techniques.

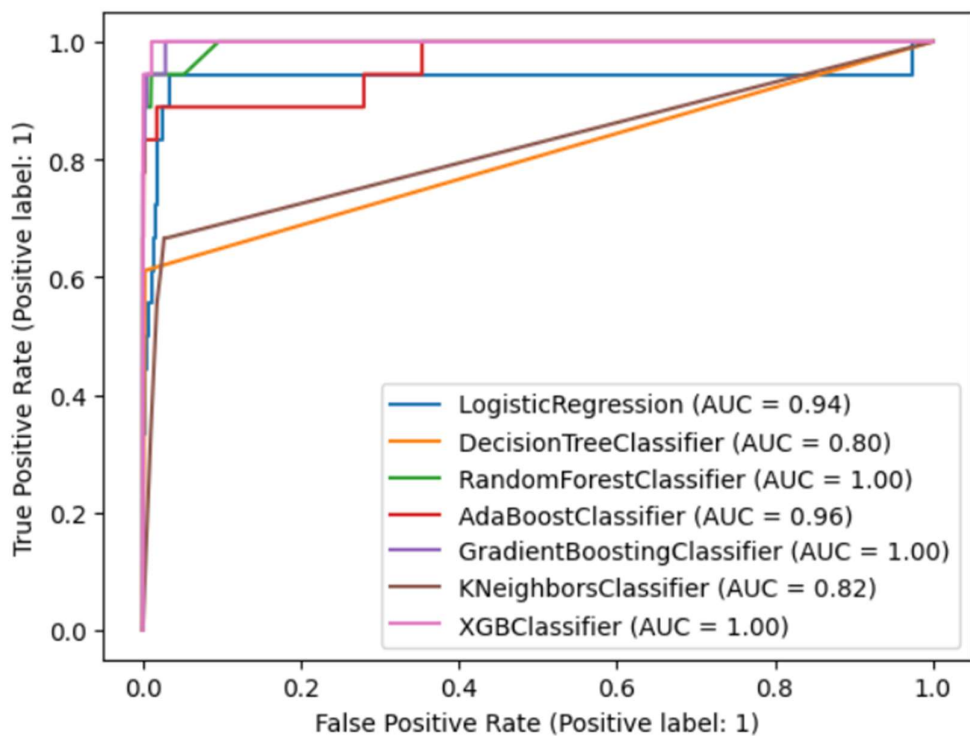
Results:

Balanced Data Results:

	Algorithm	Accuracy	
0	Logistic Regression	91.980128	
1	Decision Tree	99.656967	
2	Random Forest	99.881713	
3	Bagging	73.870357	
4	Adaboost	98.284836	
5	GradientBoosting	98.982730	
6	XGBoost	99.940856	
7	Naive-Bayes	99.716111	
8	K-nearestNeighbor	97.646085	



<Figure size 2000x600 with 0 Axes>



<Figure size 2500x2500 with 0 Axes>

Imbalanced data Results:

1. **Logistic Regression:** Testing Accuracy: 0.999655016329751
2. **Decision Tree:** Testing Accuracy: 0.9996479437715909
3. **Random Forest:** Testing Accuracy :0.9995874341073332
4. **Bagging:** Testing Accuracy :0.9997013808776888
5. **Adaboost:** Testing Accuracy: 0.9991332187055019
6. **XGBoost:** Testing Accuracy: 0.9995874341073332
7. **K-nearestNeighbor:** Testing Accuracy :0.9997013808776888