



standard
chartered

STANDARD CHARTERED GBS DIVERSITY HACKATHON -2024

FRAUD DETECTION MODEL USING AI

TEAM -02:

RAKSHITA UPADHYAY

HARINI JAYKUMAR

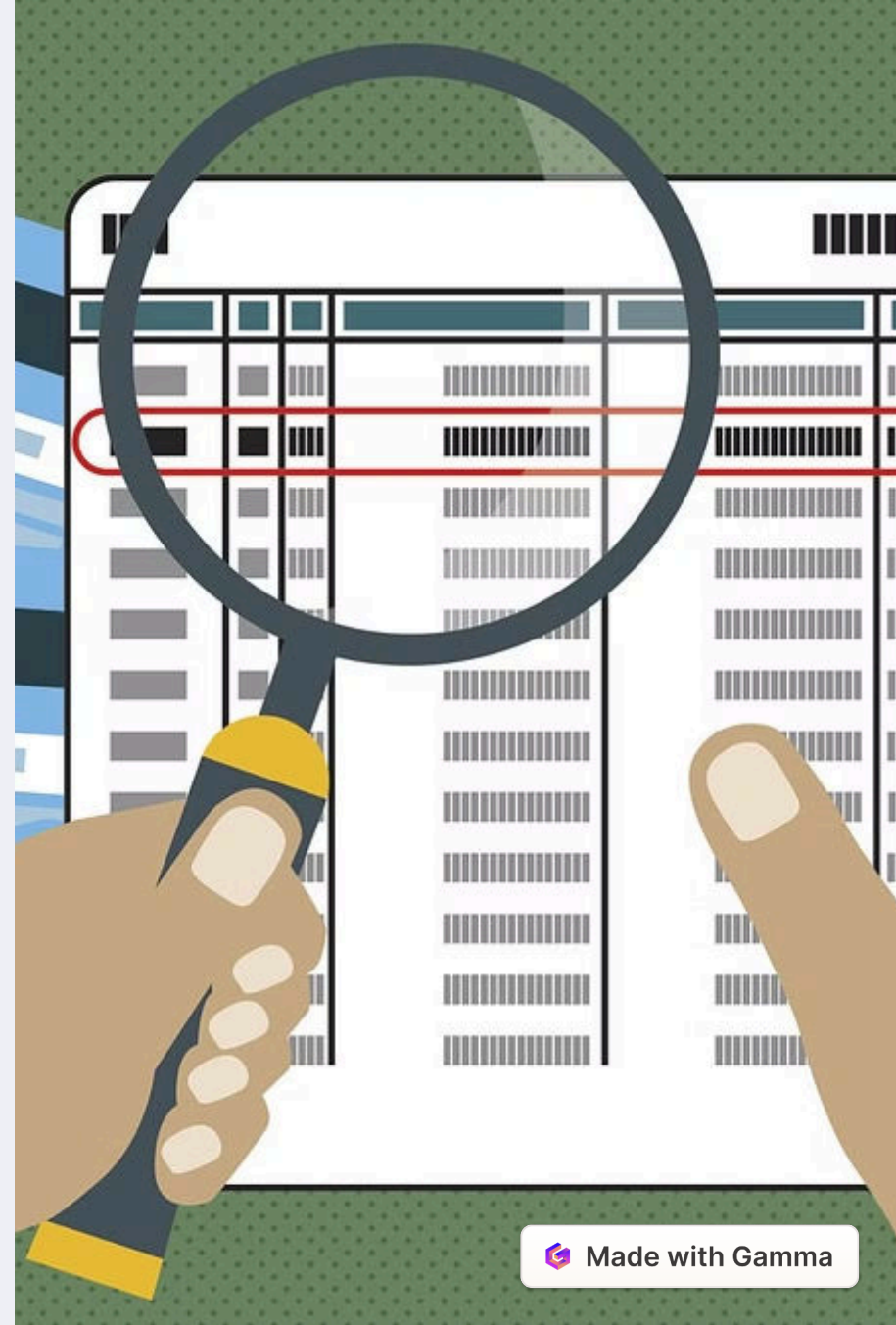
MANVI HARITWAL

SUPRABHA DESAI BOJJA

UDDARAJU VIJAYA LAKSHMI ANANYA

Fraud Detection model using AI

In today's digital age, fraudsters are becoming more sophisticated in their methods. Our Fraud Detection model leverages the power of Artificial Intelligence to detect and prevent fraudulent activities in real-time. By analyzing large amounts of data and identifying patterns, our AI-powered solution provides accurate and timely fraud alerts, helping businesses protect themselves and their customers from financial losses.



Common types of financial fraud

Identity Theft

A fraudulent acquisition and use of an individual's personal information for financial gain.

Insurance Fraud

Falsifying or exaggerating insurance claims to obtain undeserved funds.

Embezzlement

Illegal misappropriation of funds by a person entrusted to manage or monitor them.

Credit Card Fraud

Unauthorized use of someone else's credit card information to make purchases.

The impact of financial fraud on businesses and individuals



Financial Loss

Businesses and individuals suffer significant financial losses due to fraud, leading to reduced resources and opportunities for growth.



Reputation Damage

Fraud can tarnish the reputation of businesses and individuals, impacting trust and credibility in the market.



Emotional Distress

Victims of financial fraud experience emotional distress, anxiety, and stress, affecting their well-being and mental health.



Legal Consequences

Financial fraud can result in legal battles, litigation costs, and damage to personal and professional relationships.

Limitations of traditional methods

Lack of Real-time Monitoring

Traditional methods often rely on batch processing, leading to delayed detection of fraudulent activities.

Difficulty in Adapting to New Schemes

Traditional methods struggle to keep up with the evolving tactics and strategies of fraudsters.

High False Positive Rates

Manual review processes can result in a high volume of false positives, leading to wasted resources.

Machine learning and artificial intelligence in fraud detection



Data Analysis

AI processes large volumes of data to identify patterns and anomalies.



Predictive Modeling

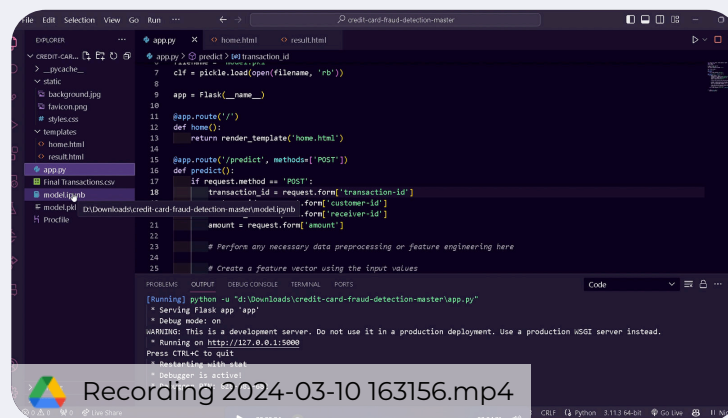
Machine learning algorithms forecast potential fraudulent activities based on historical data.




Real-time Monitoring

Smart technology enables continuous monitoring for immediate detection of suspicious transactions.

Demo Video of the Project:



 Google Docs
payshield.mp4



Technologies Incorporated:

Framework : Python Flask

Front-End : HTML , CSS , JS

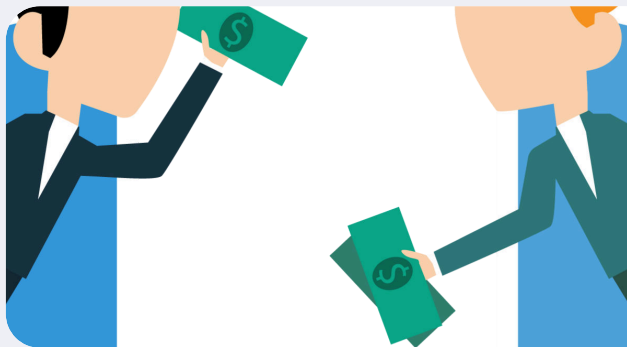
ML Prediction Model : Random Forest Method (Python)

(Compared 3 different model - Logistic Regression, XGBoost and Random Forest wherein Random forest gave the best accuracy)

Training and Testing : Google Colab

(<https://colab.research.google.com/drive/1mnoKEN5wmvf0uZxx2guhtPKAUTuVSiPW?usp=sharing>)

Dataset : Kaggle



 www.kaggle.com

Fraudulent Transaction Detection

Financial transactions labeled as fraudulent or legitimate



DATASET:

The dataset consists of 1.75 million transactions made by considering simulated users through various terminals throughout the period from January 2023 to June 2023. It marks the transaction with high amount of money and multiple transactions by the same user within a small timeframe as fraudulent.

sNo	Transaction_ID	DateTime	Customer_ID	Receiver_ID	Amount	Time_seconds	Time_Days	isFraud
1	1	01-01-2023 00:02	4961	3412	808.56	130	0	0
2	2	01-01-2023 00:07	2	1365	1442.94	476	0	1
3	3	01-01-2023 00:09	4128	8737	620.65	569	0	0
4	4	01-01-2023 00:10	927	9906	490.66	634	0	0
5	5	01-01-2023 00:10	568	8803	401.17	645	0	0
6	6	01-01-2023 00:11	2803	5490	938.54	690	0	0
7	7	01-01-2023 00:11	4684	2486	206.53	704	0	0
8	8	01-01-2023 00:11	4128	8354	253.47	713	0	0
9	9	01-01-2023 00:13	541	6212	555.63	824	0	0
10	10	01-01-2023 00:16	4554	2198	575.43	1019	0	0
11	11	01-01-2023 00:17	2000	7997	651.42	1064	0	0
12	12	01-01-2023 00:18	1948	3372	537.04	1081	0	0
13	13	01-01-2023 00:19	2938	1516	197.46	1162	0	0
14	14	01-01-2023 00:19	2989	4111	207.87	1189	0	0
15	15	01-01-2023 00:20	3842	1693	220.9	1203	0	0
16	16	01-01-2023 00:20	4361	4322	916.4	1225	0	0
17	17	01-01-2023 00:20	4177	6270	617.88	1240	0	0
18	18	01-01-2023 00:20	3700	471	845.86	1248	0	0
19	19	01-01-2023 00:21	3671	4223	637.92	1266	0	0
20	20	01-01-2023 00:21	1270	931	1378.62	1269	0	1
21	21	01-01-2023 00:22	2899	6019	353.13	1334	0	0
22	22	01-01-2023 00:23	4582	7998	476.31	1412	0	0
23	23	01-01-2023 00:25	508	9687	1376.93	1524	0	1
24	24	01-01-2023 00:26	2323	1257	628.38	1597	0	0
25	25	01-01-2023 00:26	1152	8621	528.43	1614	0	0
26	26	01-01-2023 00:27	2389	9739	306.4	1622	0	0

Features and Labels

Features:

- Transaction_id
- Customer_id
- Receiver_id
- Amount

Labels / Class variable:

- isFraud

- * It takes up the value of either 0 (or) 1.
- * 0 indicating not fraudulent and 1 indicating fraudulent transaction.

sNo	Transaction_ID	DateTime	Customer_ID	Receiver_ID	Amount	Time_seconds	Time_Days	isFraud
1	1	01-01-2023 00:02	4961	3412	808.56	130	0	0
2	2	01-01-2023 00:07	2	1365	1442.94	476	0	1

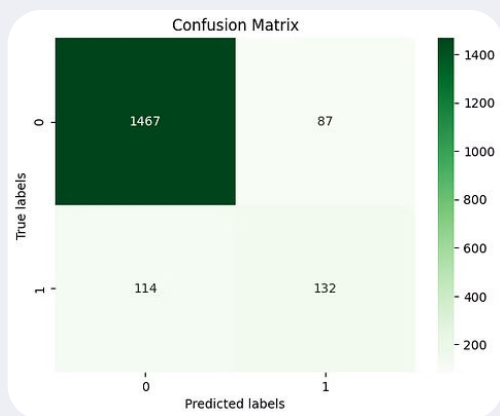
Use- Case of the Model

The prediction model built is used to detect an anomaly change in data pattern i.e whenever there's a relatively high amount transfer it classifies as fraudulent transaction otherwise non-fraudulent.

20	20	01-01-2023 00:21	1270	931	1378.62	1269	0	1
21	21	01-01-2023 00:22	2899	6019	353.13	1334	0	0
22	22	01-01-2023 00:23	4582	7998	476.31	1412	0	0
23	23	01-01-2023 00:25	508	9687	1376.93	1524	0	1

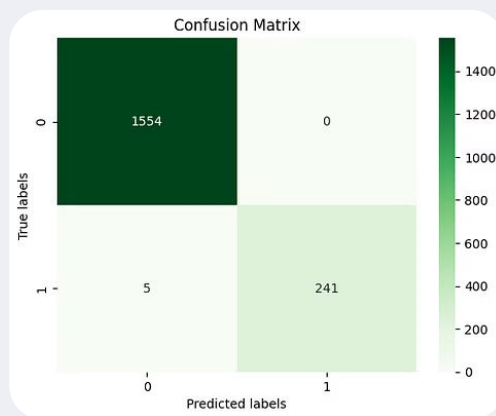
Here, in rows 20 and 23 , there's a significant change in amount feature than rows 21, 22 ; Thus the former are classified as fraudulent.

Comparing the Machine Learning Algorithms



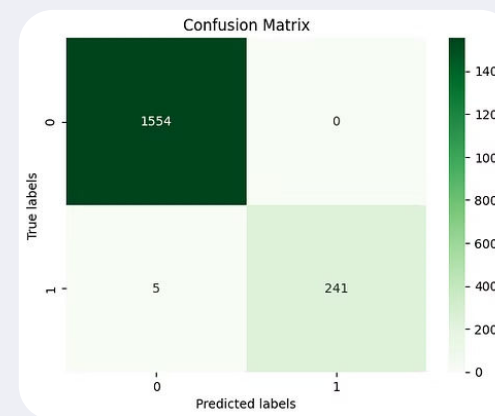
Logistic Regression

Accuracy: 88%



Random Forest

99.72%



XGBoost

99.55%

Logistical Regression

```
model_log = LogisticRegression(multi_class='multinomial', solver='lbfgs', max_iter=400 )
model_log.fit(x_train, y_train)

y_pred = model_log.predict(x_test)
mean_absolute_error(y_test, y_pred)
confusion = confusion_matrix(y_test, y_pred)
acc = accuracy_score(y_test, y_pred)
print('Accuracy Score:',int(acc*100),'%')
```

Accuracy Score: 89 %

Random Forest Classifier

```
model_rf = RandomForestClassifier(n_estimators = 400)
model_rf.fit(x_train , y_train)

pred = model_rf.predict(x_test)
confusion = confusion_matrix(y_test, pred)
acc = accuracy_score(y_test, pred)
print('Accuracy Score:',(acc*100),'%')
```

Accuracy Score: 99.33333333333333 %

XGB Classifier

```
model_xgb = xgb.XGBClassifier()
model_xgb.fit(x_train , y_train)

pred = model_xgb.predict(x_test)
confusion = confusion_matrix(y_test, pred)
acc = accuracy_score(y_test, pred)
print('Accuracy Score:',(acc*100),'%')
```

Accuracy Score: 99.22222222222223 %

Why Random Forest?

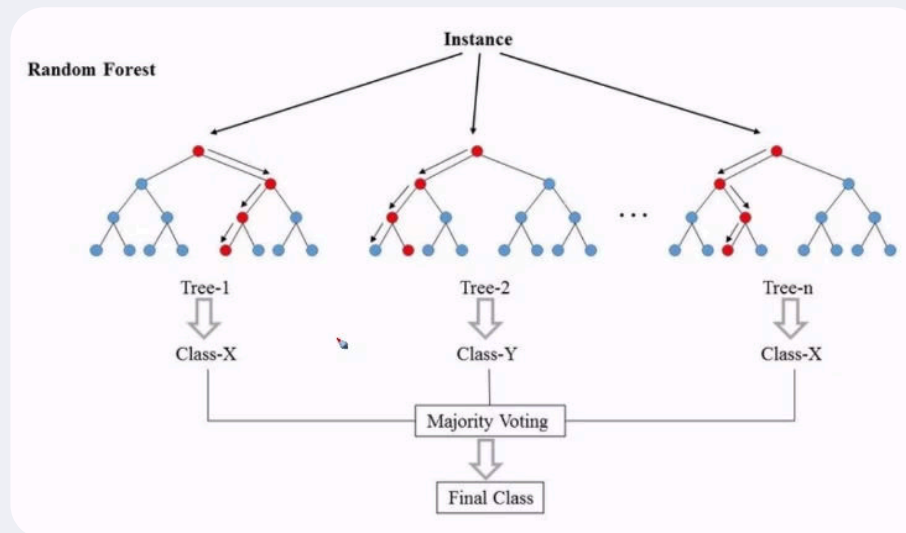
1. **Accuracy:** The random forest model achieved a high accuracy score of 99.72%, indicating strong predictive performance on the dataset.
2. **Ensemble Advantage:** Random forests have the advantage of being an ensemble model, combining the predictions of multiple decision trees. This often leads to improved generalization and robustness compared to individual models like logistic regression.
3. **Overfitting Mitigation:** Random forests are less prone to overfitting compared to complex models like XGBoost. This is evident from the accuracy scores, where the random forest outperformed XGBoost, suggesting that the former may have a better balance between bias and variance.
4. **Confusion Matrix Comparison:** The confusion matrix provides insights into the model's performance, especially in terms of precision, recall, and F1 score for each class. If the random forest also exhibited better performance in terms of these metrics, it would further support its selection.

Random Forest - Machine learning Approach

Random Forest is an ensemble learning method used in machine learning for both classification and regression tasks. It constructs a multitude of decision trees during training and outputs the mode (for classification) or average prediction (for regression) of the individual trees.

The key idea is to introduce randomness at two levels: by using a random subset of the training data to build each tree (bagging), and by considering only a random subset of features at each split in the decision tree. This randomness helps prevent overfitting and improves the model's robustness and accuracy.

The final prediction is a combination of the predictions from all the individual trees, resulting in a powerful and versatile algorithm suitable for a wide range of applications.



Conclusion and Future

The fraud detection model has high accuracy and effectively identifies fraud, enhancing security. Continuous monitoring and updates are needed to keep up with evolving fraud patterns. This model is a strong foundation for securing financial transactions, but ongoing vigilance and improvements are necessary to stay ahead of emerging threats.

Future efforts should improve the model's ability to detect new fraud methods by training it regularly with up-to-date data. Implementing advanced anomaly detection techniques and real-time monitoring can strengthen the system even more. Working with industry experts and incorporating cutting-edge technologies like deep learning or blockchain can add extra security layers. Moreover, enhancing interpretability, trust, and expanding the model's use across different industries are important research directions for reinforcing fraud detection methods.

