Credit Card Fraud Detection and Rectification- A Webapp

DR. SHUDHODHAN BOKEFODE[1], HARSHALI KADAM[2], NAWAF LOKARE[3], ADITYA BHURAN[4], GAURAV NATEKAR[5]

[1] *Department of Computer EngineeringTerna Engineering College, Navi Mumbai, India*

[2, 3, 4, 5] *Student, Department of Computer EngineeringTerna Engineering College, Navi Mumbai, India*

Abstract— The Credit Card Fraud Detection Web App is an advanced online platform designed to combat the growing threat offraudulent activities related to Credit Card in financial transactions. Leveraging cutting-edge Artificial Intelligence (AI) algorithms and machine learning techniques we detect the transaction is Fraud or not The web application provides a robust and proactive solutionto identify, prevent and also rectify the unauthorized credit cardusage in real-time. We use a variety of machine learning models/classifiers to accomplish this, such as Random Forest (RF), Logistic Regression (LR), Naive Baiye(NB)

Index Terms— Random Forest (RF), Logistic Regression (LR), Naive Baiye(NB)

I. INTRODUCTION

The Credit Card Fraud Detection Web App represents a critical tool in the fight against credit card fraud, minimizing financial losses for both consumers and financial institutions.

By harnessing the power of AI and real-time monitoring, this platform sets a new standard for secure and trustworthy financial transactions in the digital age.

The Credit Card Fraud Detection Web App aims to address theescalating issue of credit card fraud, which poses a significantthreat to consumers

With the increasing reliance on digital transactions, fraudulent activities have become more sophisticated, resulting in substantial financial losses and eroding trust in online payment systems.

II. PROBLEM STATEMENT

The problem statement encompasses creating robust algorithms that can analyse vast amounts of transaction data, detecting patterns indicative of fraudulent activity.

The goal is to minimize false positives, ensuring legitimate transactions are not wrongly flagged, while maintaining a high level of accuracy in identifying suspicious or anomalousbehaviour.

This involves a constant adaptation to evolving fraud tactics and a balance between sensitivity and specificity in the detection process.

Additionally, the problem involves addressing challenges such as imbalanced datasets, where instances of fraud are relatively rare compared to legitimate transactions.

Solutions need to be resilient to evolving fraud techniques, adaptable to new patterns, and capable of providing timely alerts or interventions to prevent unauthorized transactions. Integration with advanced technologies like machine learning and artificial intelligence plays a crucial role in developing efficient and scalable credit card fraud detection systems.

III. PROPOSED METHODOLOGY

A Credit card fraud web app where the user has to enter his/herpersonal details and based on the transaction our system will apply various ML classifiers on the data

And the transaction is fraudalent or not would be reported andnotified

Additional feature for the members would be that they can justprovide the credit card name and number and our system would take their weekly bank statement and always keep them in checkif fraud is detected or not

If fraud is detected our web app will be the middle

agent between the user and the RBI

And the user will get the money back in 30 days in his comfort space without checking or reporting fraud

Everything will be reported to user via EMAIL and WATSAPP about the fraud detection and rectification of false transaction Such a system typically involves the following components

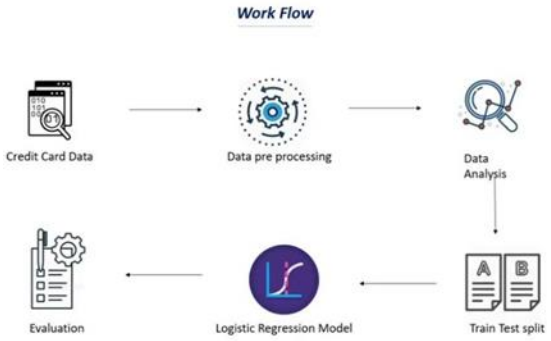
IV. FLOW OF PROJECT

The five phases are as follows:

1) Collection of Datasets
2) Pre-processing (Noise Removal)
3) Data Analysis
4) Train and Test data
5) Applied Machine Learning classifiers

Evaluation

Flow of the Proposed System

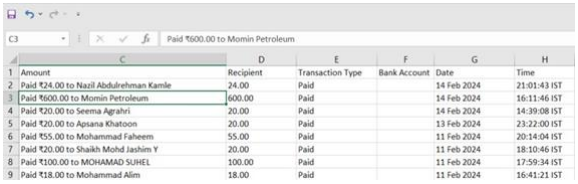


As demonstrated in the figure, the methodology to extract the output contains the several steps that are described below:

1. Data Collection:

Credit card fraud detection involves collecting and analyzing various data points, including transaction history, user behavior, and geographical information. Machine learning algorithms canthen identify patterns and anomalies to flag potential fraudulent activities. In addition to transaction history, user behavior, and geographical information, credit card fraud detection systems often consider factors like purchase frequency, transaction amount, device used, IP address, time of day, and merchant reputation. This comprehensive data approach enhances the accuracy of fraud detection algorithms.

Here we took real-time data of G-PAY and converted it into useful data

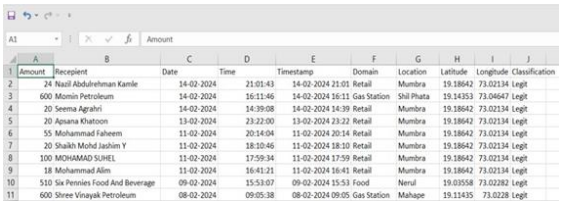


	C	D	E	F	G	H
1	Amount	Recipient	Transaction Type	Bank Account	Date	Time
2	Paid ₹24.00 to Nazil Abdulrehman Kamle	24.00	Paid		14 Feb 2024	21:01:43 IST
3	Paid ₹600.00 to Momin Petroleum	600.00	Paid		14 Feb 2024	16:11:46 IST
4	Paid ₹20.00 to Seema Agrahri	20.00	Paid		14 Feb 2024	14:39:08 IST
5	Paid ₹20.00 to Apsana Khatoon	20.00	Paid		13 Feb 2024	23:22:00 IST
6	Paid ₹55.00 to Mohammad Faheem	55.00	Paid		11 Feb 2024	20:14:04 IST
7	Paid ₹20.00 to Shaikh Mohd Jashim Y	20.00	Paid		11 Feb 2024	18:10:46 IST
8	Paid ₹100.00 to MOHAMAD SUHEL	100.00	Paid		11 Feb 2024	17:59:34 IST
9	Paid ₹18.00 to Mohammad Alim	18.00	Paid		11 Feb 2024	16:41:21 IST

2. Pre-processing:

Data preprocessing in credit card fraud detection involves steps like handling missing values, scaling numerical features,encoding categorical variables, and addressing class imbalance. Additionally, it may include normalization, feature engineering,and outlier detection to ensure the input data is suitable for training robust machine learning models

As we got raw data from GPAY we performed data cleaning ,data cleaning for credit card fraud detection may involve removing duplicate records, dealing with noisy data, and standardizing features. Feature selection techniques can be applied to identify and retain the most relevant attributes



	A	B	C	D	E	F	G	H	I	J
1	Amount	Recepient	Date	Time	Timestamp	Domain	Location	Latitude	Longitude	Classification
2	24	Nazil Abdulrehman Kamle	14-02-2024	21:01:43	14-02-2024 21:01	Retail	Mumbra	19.18642	73.02134	Legit
3	600	Momin Petroleum	14-02-2024	16:11:46	14-02-2024 16:11	Gas Station	Shil Phata	19.14353	73.04647	Legit
4	20	Seema Agrahri	14-02-2024	14:39:08	14-02-2024 14:39	Retail	Mumbra	19.18642	73.02134	Legit
5	20	Apsana Khatoon	13-02-2024	23:22:00	13-02-2024 23:22	Retail	Mumbra	19.18642	73.02134	Legit
6	55	Mohammad Faheem	11-02-2024	20:14:04	11-02-2024 20:14	Retail	Mumbra	19.18642	73.02134	Legit
7	20	Shaikh Mohd Jashim Y	11-02-2024	18:10:46	11-02-2024 18:10	Retail	Mumbra	19.18642	73.02134	Legit
8	100	MOHAMAD SUHEL	11-02-2024	17:59:34	11-02-2024 17:59	Retail	Mumbra	19.18642	73.02134	Legit
9	18	Mohammad Alim	11-02-2024	16:41:21	11-02-2024 16:41	Retail	Mumbra	19.18642	73.02134	Legit
10	510	Six Pennies Food And Beverage	09-02-2024	15:53:07	09-02-2024 15:53	Food	Nerul	19.03558	73.02282	Legit
11	600	Shree Vinayak Petroleum	08-02-2024	09:05:38	08-02-2024 09:05	Gas Station	Mahape	19.11435	73.0228	Legit

3. Data Analysis:

Data analysis for credit card fraud detection involves exploring patterns, trends, and anomalies within the collected data. Descriptive statistics, visualizations, and correlation analysis help uncover insights. Advanced techniques, such as clustering or anomaly detection algorithms, can identify unusual patterns that may indicate fraudulent activities. This analysis informs the development of accurate models for detecting and preventing credit card fraud.

4. Train and Test data

For credit card fraud detection, you typically split your datasetinto training and testing sets. Use a majority for training (e.g., 80%) and the rest for testing. Ensure

both sets have a balancedrepresentation of normal and fraudulent transactions to avoid biased models.

5. Applied Machine Learning classifiers

In this proposed system applied different Machine Learning classifiers like Random Forest (RF), Logistic Regression (LR), Naive Baiye(NB)

a. Logistic Regression (LR):

Logistic regression is a statistical method for binary classification. It models the probability of an event occurring as a function of input features. By applying a logistic (S- shaped) curve, it maps a linear combination of input variablesto a range between 0 and 1, making it suitable for estimating probabilities and making binary decisions.

Fig. 3. Result of LR

```
===== LogisticRegression =====

Cross Validation Mean Score:  87.4%

Model Accuracy:  90.10000000000001%

Confusion Matrix:
 [[174116  16361]
 [ 21442 169029]]

Classification Report:
              precision    recall  f1-score   support

           0       0.89      0.91      0.90    190477
           1       0.91      0.89      0.90    190471

    accuracy                           0.90    380948
   macro avg       0.90      0.90      0.90    380948
weighted avg       0.90      0.90      0.90    380948
```

b. Naïve Baiyes classifier (NB):

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.Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier

```
===== Naive Baiye Classifier =====

Cross Validation Mean Score:  83.7%

Model Accuracy:  84.8%

Confusion Matrix:
 [[170311  20166]
 [ 37647 152824]]

Classification Report:
              precision    recall  f1-score   support

           0       0.82      0.89      0.85    190477
           1       0.88      0.80      0.84    190471

    accuracy                           0.85    380948
   macro avg       0.85      0.85      0.85    380948
weighted avg       0.85      0.85      0.85    380948
```

Fig. 4. Result of NB

c. Random Forest (RF):

Random Forest is an ensemble machine learning algorithm. It builds multiple decision trees during training and combines their predictions for more accurate and robust results. Itmitigates overfitting and increases accuracy by using features randomly. It's widely used for classification and regressiontasks, offering strong performance and interpretability.

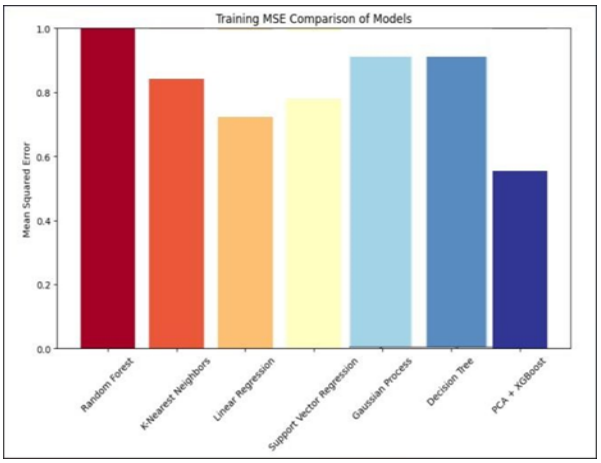
```
========================= Model Evaluation Results ======================

===== RandomForest Classifier =====

Cross Validation Mean Score:  93.0%

Model Accuracy:  100.0%

Confusion Matrix:
 [[190477      0]
 [     0 190471]]

Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00    190477
           1       1.00      1.00      1.00    190471

    accuracy                           1.00    380948
   macro avg       1.00      1.00      1.00    380948
weighted avg       1.00      1.00      1.00    380948
```

Fig. 4. Result of RF

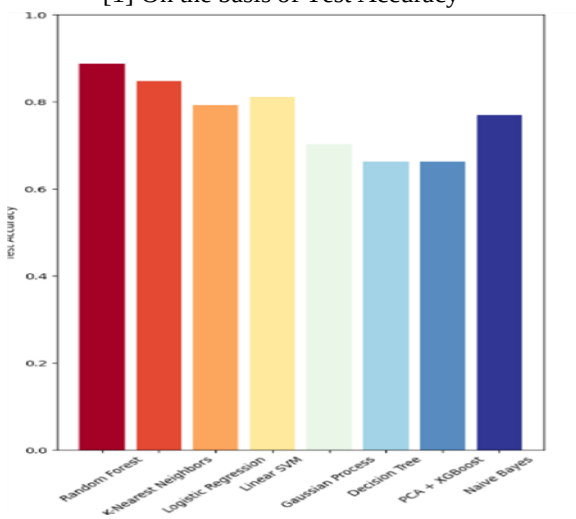
6. RESULT AND ANALYSIS

Algorithm	ccuracyscore	Precision score
RandomForest	100.00%	1.00
Logistic Regression	90.100%	0.89
Naïve Baiyes	84.8%	0.82

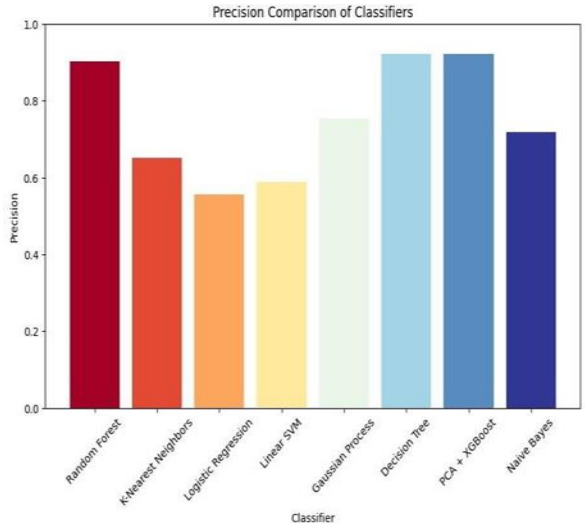
During our analysis of various machine learning classifiers, we witnessed outstanding performance measure and its comparisons



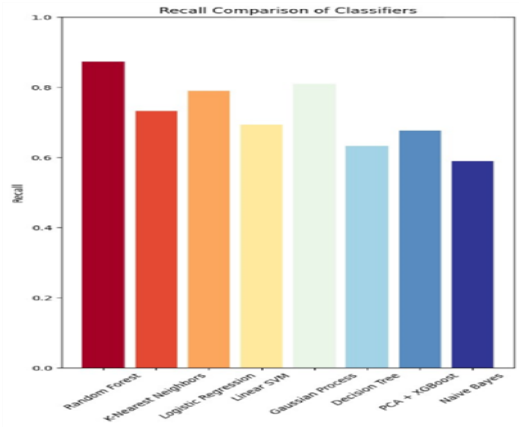
[1] On the basis of Test Accuracy –



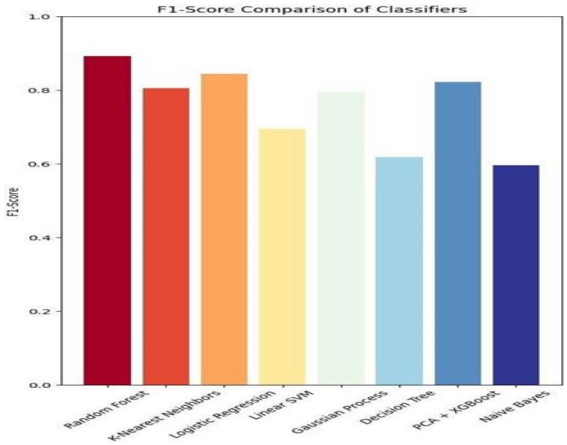
[2] On the basis of precision –



[3] On the basis of Recall-



[4] On the basis of F-1 score-



7. To remove the outliers Isolation Forest

Isolation Forest is a technique for identifying outliers in data. Any data point/observation that deviates significantly from the other observations is called an Anomaly/Outlier. Isolation Forests(IF), similar to Random Forests, are build based on decision trees. And since there are no pre-defined labels here,it is an unsupervised model.

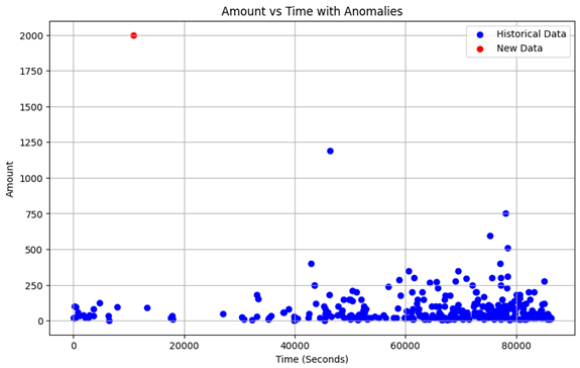
How the Isolation Forest Algorithm Works

The Isolation Forest Algorithm takes advantage of the following properties of anomalous samples (often referred to as outliers):

Fewness — anomalous samples are a minority and there will only be a few of them in any dataset.

Different — anomalous samples have values/attributes that are very different from those of normal samples.

These two properties make it easier to isolate anomaloussamples from the rest of the data in comparisonto normal points.



Notice how in the figure above, we can isolate an anomalous point from the rest of the data with just one line and can be visually represented too

CONCLUSION

In our credit card fraud detection project, we implemented various machine learning algorithms to assess their performance. Remarkably, the Random Forest algorithm exhibited outstanding accuracy, achieving a perfect 100%. Logistic Regression also demonstrated strong performance with an accuracy of 90.1%. Naive Bayes, while effective, displayed a slightly lower accuracy at 84.8%.

These results suggest that our models are generally effective in identifying fraudulent transactions. However we considered essential additional metrics such as precision,recall, and F1-score to gain a more comprehensive understanding of each model's performance. Furthermore, exploring techniques like hyperparameter tuning, feature engineering, and ensemble methods may offer opportunitiesto enhance the overall effectiveness of our fraud detection system. Continuous anamoly detection and refinement and optimization of the models can contribute to even greater accuracy and robustness in identifying fraudulent activities.

REFERENCES

[1] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim and A. K. Nandi, "Credit Card Fraud Detection Using AdaBoost and Majority Voting", IEEE, vol. 6, pp. 14277-14284, 2018.

[2] D. Tanouz, R. R. Subramanian and D. Eswar, "Credit Card Fraud Detection Using Machine Learning", IEEE, 2021.

[3] E. Ileberi, Y. Sun and Z. Wang, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost", IEEE, vol. 9, no. 5, pp. 165286-165294, 2021.

[4] S. Khatri, A. Arora and A. P. Agrawal, "Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison", IEEE, 2020.

[5] J. O. Awoyemi and A. O, "Credit card fraud detection using machine learning techniques: A comparative analysis", IEEE, 2017.

[6] M. K. Dejan Varmedja, "Credit Card Fraud Detection - Machine Learning methods", IEEE Xplore, 2019.

[7] G. L. S. Xuan, "Random forest for credit card fraud detection", IEEE 15th International Conference on Networking Sensing and Control (ICNSC), 2018.