IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Special Issue 2, April 2023

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | SJIF: 8.423|| Volume 12, Special Issue 2, April 2023 ||

8th National Conference on Advanced Computing Technologies (NCACT'23)

Organized by

Department of CSE, Velammal Engineering College, Chennnai, India

Credit Card Fraud Detection using Data Science Technique

Jagadesh Subramaniam K.U [1], Rohit Jha D [2], Karthick D [3], Gokul M [4]

U.G. Student, Department of Information Technology, GSBT Engineering College, Chennai, Tamil Nadu, India[1]

U.G. Student, Department of Information Technology, GSBT Engineering College, Chennai, Tamil Nadu, India[2]

U.G. Student, Department of Information Technology, GSBT Engineering College, Chennai, Tamil Nadu, India[3]

U.G. Student, Department of Information Technology, GSBT Engineering College, Chennai, Tamil Nadu, India[3]

ABSTRACT: The identification of fraudulent credit card transactions is crucial to ensure that customers are not charged for purchases they did not make. Data Science and Machine Learning are essential tools in addressing this issue. This project aims to demonstrate how machine learning can be used to model a dataset for Credit Card Fraud Detection. This involves analyzing past credit card transactions, particularly those that were fraudulent, to build a model that can recognize whether a new transaction is fraudulent or not. The goal is to detect all fraudulent transactions while minimizing false alarms. The project includes analyzing and pre-processing datasets and applying anomaly detection algorithms such as Local Outlier Factor and Random Forest algorithm to the PCA-transformed Credit Card Transaction data.

KEYWORDS*:* Credit card fraud, Applications of Machine Learning, Data Science, Random Forest algorithm, Automated Fraud Detection

I.INTRODUCTION

Credit card fraud occurs when someone uses another person's credit card without their knowledge or consent for personal reasons, and the owner and card issuer are unaware of the activity. Detecting and preventing fraud involves monitoring the behavior of user populations to identify objectionable actions, including fraud, intrusion, and defaulting. This problem requires the attention of communities such as machine learning and data science, where automation can help solve the problem. However, there are various challenges in implementing a fraud detection system, including class imbalance and changes in transaction patterns over time. To address this issue, machine learning algorithms analyze authorized transactions and flag suspicious ones for investigation by professionals. The feedback from these investigations is then used to train and update the algorithm to improve fraud detection performance over time. This process helps prevent fraudulent activity and protect against similar incidents in the future.

II. RELATED WORK

Fraud refers to a deliberate act that aims to attain financial or personal benefit by breaking the law, rules, or policies. Several techniques have been used in fraud detection, such as data mining, automated fraud detection, and adversarial detection. However, these methods have not provided a permanent and consistent solution. Researchers have used outlier mining, outlier detection mining, and distance sum algorithms to detect fraudulent transactions accurately. Unconventional techniques such as hybrid data mining/complex network classification algorithm and artificial genetic algorithm have also been employed. The latter approach proved effective in identifying fraudulent transactions while minimizing false alerts, but it faced a classification problem with variable misclassification costs. Efforts have also been made to improve the alert-feedback interaction in the case

of fraudulent transactions, where the authorized system would be alerted, and feedback would be sent to deny the ongoing transaction.

Criminals are constantly adapting their fraudulent strategies, which requires the development of new fraud detection methods. There are various types of fraud, including

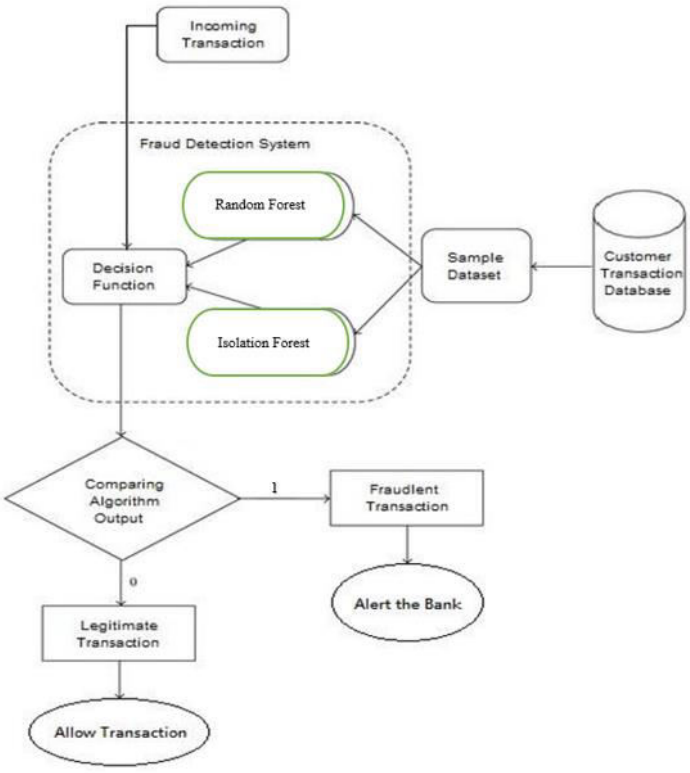
- credit card fraud (both online and offline),
- card theft,
- account bankruptcy,
- device intrusion, application fraud,
- counterfeit card, and
- telecommunication fraud.

To combat these fraudulent activities, several approaches are used, including

- Artificial neural networks,
- Fuzzy logic,
- Decision trees,
- Support vector machines,
- Bayesian networks,
- Hidden Markov models, and
- K-nearest neighbor.

These methods are used to detect fraudulent activities and protect against financial losses.

The professionals who investigate the suspicious transactions provide feedback to the automated system, which is then utilized to train and update the algorithm, with the ultimate goal of enhancing the system's ability to detect fraud over time.



Flow diagram to detect fraudulent transaction

|| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | SJIF: 8.423|| Volume 12, Special Issue 2, April 2023 ||

8th National Conference on Advanced Computing Technologies (NCACT'23)

Organized by

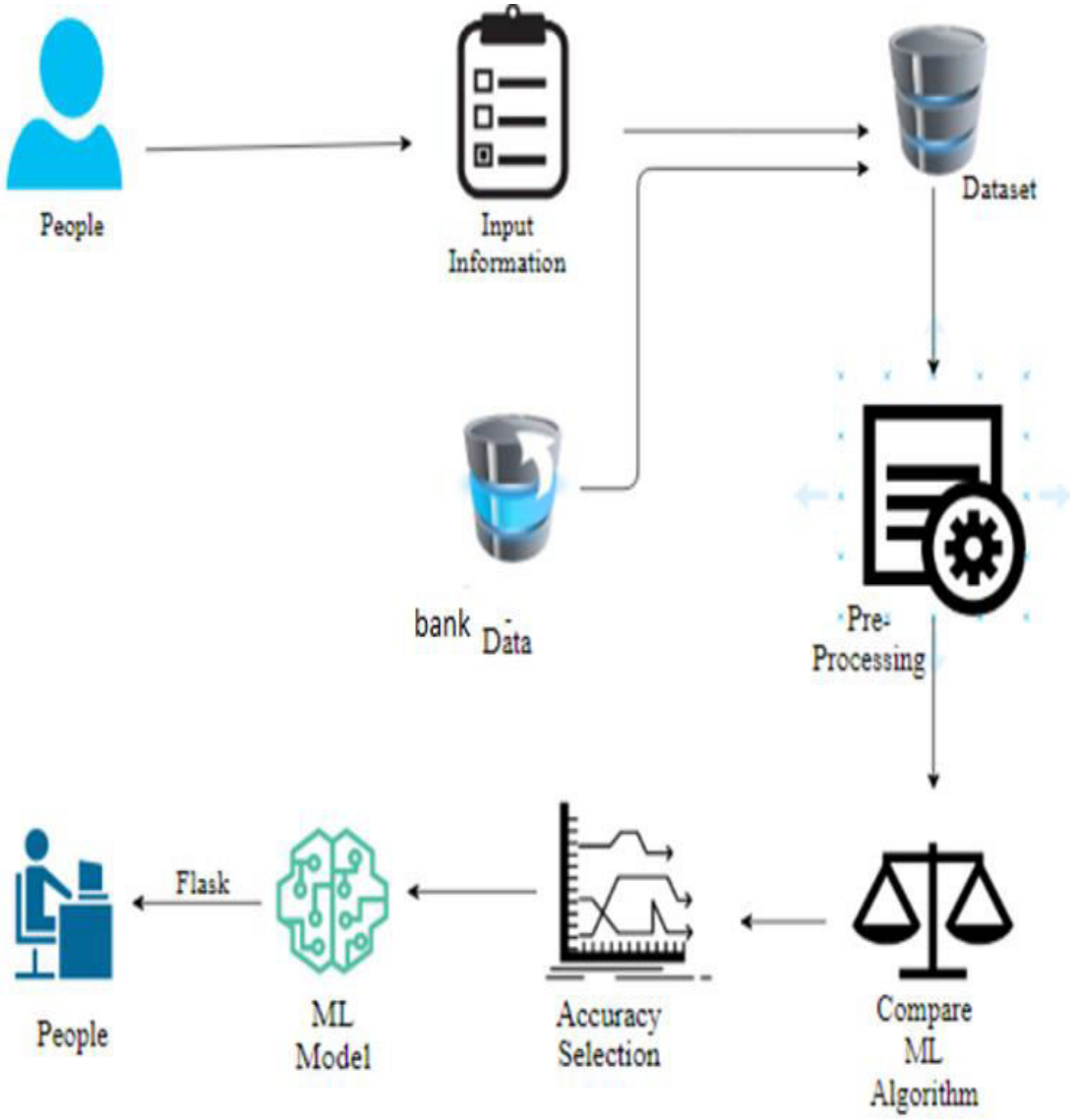
Department of CSE, Velammal Engineering College, Chennnai, India

III. METHODOLOGY

First of all, we obtained our dataset from Kaggle, a data analysis website which provides datasets.Inside this dataset, there are 31 columns out of which 28 are named as v1-v28 to protect sensitive data.

The other columns represent Time, Amount and Class. Time shows the time gap between the first transaction and the following one. Amount is the amount of money transacted. Class 0 represents a valid transaction and 1 represents a fraudulent one.

After checking this dataset, we plot a histogram for every column. This is done to get a graphical representation of the dataset which can be used to verify that there are no missing any values in the dataset. This is done to ensure that we don't require any missing value imputation and the machine learning algorithms can process the dataset smoothly.



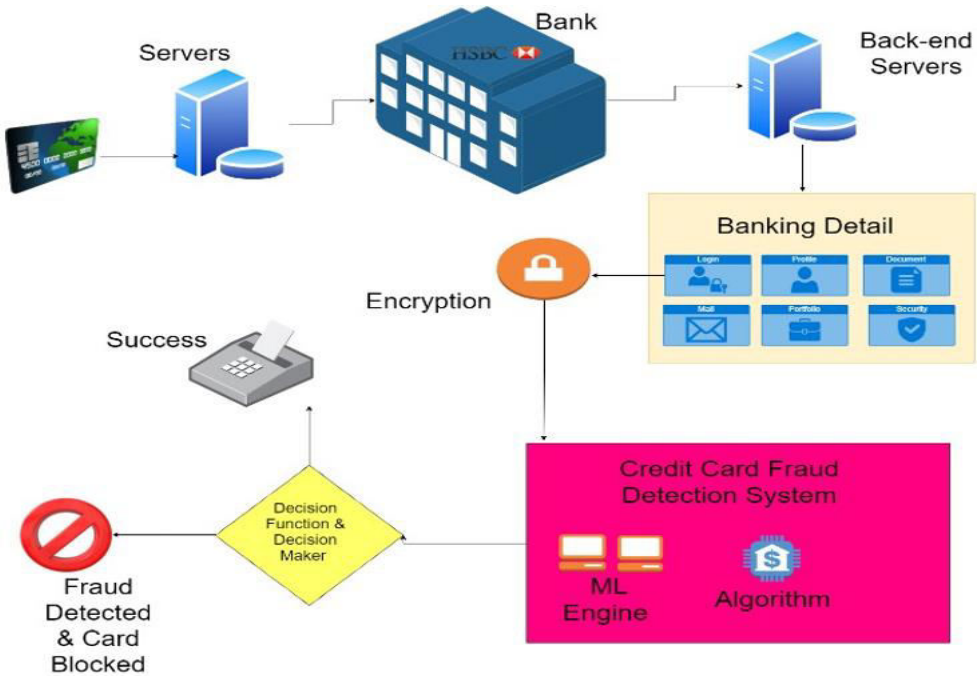
System Architecture Diagram

|| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | SJIF: 8.423|| Volume 12, Special Issue 2, April 2023 ||

8th National Conference on Advanced Computing Technologies (NCACT'23)

Organized by

Department of CSE, Velammal Engineering College, Chennnai, India



This above diagram is represent the step by step process of Credit Card Fraud Detection System Using ML algorithm **(Random Forest)**

Random Forest Algorithm

i. **Data preprocessing:** Start by collecting the data related to credit card transactions, and cleaning and preprocessing the data. This includes removing duplicates, dealing with missing values, and converting categorical variables to numerical values.
ii. **Feature selection:** Select the relevant features that can help in detecting fraud. This can be done by analyzing the correlation of each feature with the target variable (fraudulent or not).
iii. **Data splitting:** Split the dataset into training and testing sets. The training set will be used to train the Random Forest model, while the testing set will be used to evaluate the model's performance.
iv. **Model training:** Train the Random Forest model on the training set, using the selected features. Adjust the hyperparameters of the model, such as the number of trees, maximum depth of each tree, and the minimum number of samples required to split an internal node.
v. **Model evaluation:** Evaluate the performance of the Random Forest model on the testing set. Use metrics such as accuracy, precision, recall, and F1-score to evaluate the model's performance.
vi. **Model deployment:** Once the model has been trained and evaluated, it can be deployed to detect fraud in real-time. The model can be integrated with the credit card processing system to automatically flag suspicious transactions.

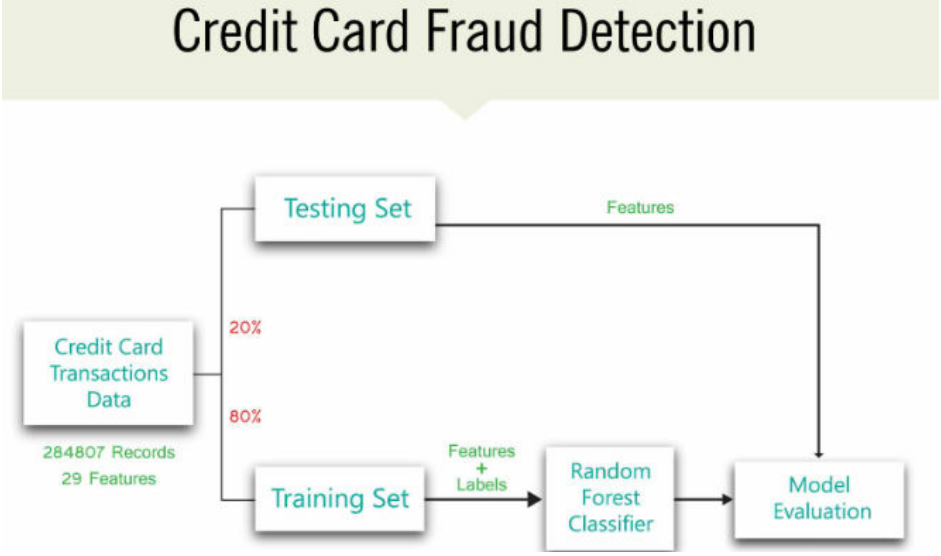
It's important to note that credit card fraud is a rare event, and the dataset is typically imbalanced. Therefore, it's important to balance the dataset to improve the model's performance.

|| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | SJIF: 8.423|| Volume 12, Special Issue 2, April 2023 ||

8th National Conference on Advanced Computing Technologies (NCACT'23)

Organized by

Department of CSE, Velammal Engineering College, Chennnai, India



IV. EXPERIMENT RESULT

The code evaluates the accuracy and precision of the algorithms used for fraud detection by comparing the number of false positives detected by the algorithms with the actual values.

Only 10% of the dataset is used to run the code faster. The results of the algorithm are printed along with the classification report for each algorithm.

In the output, class 0 represents valid transactions, while class 1 represents fraudulent transactions. The results are compared with the class values to identify false positives.

The code also uses the complete dataset to provide a comprehensive evaluation of the algorithms.

Result with complete dataset

```
Classification report of Random Forest Results:

              precision    recall  f1-score   support

           0       0.99      1.00      0.99       789
           1       0.98      0.95      0.96       134

    accuracy                           0.99       923
   macro avg       0.98      0.97      0.98       923
weighted avg       0.99      0.99      0.99       923


Confusion Matrix result of Random Forest Classifier is:
 [[786   3]
 [  7 127]]

Sensitivity :  0.9961977186311787

Specificity :  0.9477611940298507

Cross validation test results of accuracy:
[0.97723577 0.99349593 0.9804878  0.98536585 0.98211382]

Accuracy result of Random Forest Classifier is: 98.3739837398374
```

|| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | SJIF: 8.423|| Volume 12, Special Issue 2, April 2023 ||

8th National Conference on Advanced Computing Technologies (NCACT'23)

Organized by

Department of CSE, Velammal Engineering College, Chennnai, India

V. CONCLUSION

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The best accuracy on public test set is higher accuracy score will be find out. This application can help to find the Prediction of credit card fraud or not.

This article has discussed the various fraudulent methods used in credit card transactions and the techniques used to detect them. It has also reviewed recent developments in the field. Additionally, the article has explained how machine learning can be utilized to enhance fraud detection results. The article provides a detailed explanation of the machine learning algorithm used, including its pseudocode, implementation, and experimental results. Overall, the article provides a comprehensive overview of credit card fraud detection, its challenges, and the potential benefits of using machine learning algorithms for fraud detection.

REFERENCES

[1]Q. Wu, M. Zhou, Q. Zhu, Y. Xia, and J. Wen, "MOELS: Multiobjective evolutionary list scheduling for cloud workflows," IEEE Trans. Autom. Sci. Eng., vol. 17, no. 1, pp. 166–176, Jan. 2020.

[2] L. Huang, M. Zhou, and K. Hao, "Non-dominated immune-endocrine short feedback algorithm for multi-robot maritime patrolling," IEEE Trans. Intell. Transp. Syst., vol. 21, no. 1, pp. 362–373, Jan. 2020.

[3] X. Wang, K. Xing, C.-B. Yan, and M. Zhou, "A novel MOEA/D for multiobjective scheduling of flexible manufacturing systems," Complexity, vol. 2019, pp. 1–14, Jun. 2019.

[4] J. J. Liang, C. T. Yue, and B. Y. Qu, "Multimodal multi-objective optimization: A preliminary study," in Proc. IEEE Congr. Evol. Comput. (CEC), Jul. 2016, pp. 2454–2461.

[5] P. S. Oliveto, D. Sudholt, and C. Zarges, "On the benefits and risks of using fitness sharing for multimodal optimisation," Theor. Comput. Sci., vol. 773, pp. 53–70, Jun. 2019.

[6] "Credit Card Fraud Detection Based on Transaction Behaviour -by John Richard D. Kho, Larry A. Vea" published by Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8,2017

[7]"Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy" published by IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 29, NO. 8, AUGUST2018

[8]Credit Card Fraud Detection through Parenclitic Network AnalysisBy Massimiliano Zanin, Miguel Romance, ReginoCriado, and SantiagoMoral" published by Hindawi Complexity Volume 2018, Article ID 5764370

INNO SPACE
SJIF Scientific Journal Impact Factor

SJIF: 8.423

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INTERNATIONAL CENTRE

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details