International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 01 | Jan 2023 www.irjet.net p-ISSN: 2395-0072

# A Review of deep learning techniques in detection of anomaly incredit card transaction.

**Kiran Kumar M [1], Junaid Ahmed S.Y [2], Mohammed Faizan Ghani [3], Husama PM [4], Mahesh Basavaraj [5]**

*1,2,3,4 Students, Department of Computer Science & Engineering, T John Institute of Technology,Bengaluru, India*
*5Asst. Professor, Department of Computer Science & Engineering, T John Institute of Technology, Bengaluru.*

---

**Abstract -** Credit card fraud is a common occurrence that causes enormous financial losses. Online purchases have dramatically expanded, and a major portion of those purchases are made using credit cards. As a result, banks and other financial organizations fund the development of software that identify credit card fraud. Fraudulent transactions can occur in a variety of ways and fall under a number of distinct categories. Credit card firms need to identify fraudulent credit card transactions in order to avoid having their customers' accounts charged for goods they did not buy. Machine learning and data science aid in resolving these problems. The legal transactions are mixed in with the fraudulent transactions, so it is impossible to effectively identify the fraudulent transactions using simple identification approaches that compare both the fraudulent and legitimate data. With the use of credit card fraud detection, this research aims to demonstrate the modelling of a knowledge set using machine learning. Our objective is to eliminate erroneous fraud classifications while detecting 100% of fraudulent transactions. A typical categorization sample would be credit card fraud detection. On the PCA converted Credit Card Transaction data, we concentrated on analyzing and pre-processing data sets, as well as deploying numerous anomaly detection techniques such as the Local Outlier Factor and Isolation Forest algorithm, as well as one class SVM (Support Vector Machine).

**Key Words: Credit Card Fraud Detection, Support Vector Machine, Data Science, Local Outlier Factor, and Isolation Forest Algorithm**

## 1. INTRODUCTION

Groups like machine learning and data science should pay attention to this issue because it has the potential to be automatically resolved. This problem is quite challenging from a learning perspective because there are far more legitimate transactions than fraudulent ones. Additionally, over time, statistical aspects cause the transaction patterns to change often or regularly. With numerous frauds, primarily the majority of people in the world are concerned about credit card scams because they have been in the news so regularly recently. The credit card database is seriously affected when a legitimate transaction is contrasted against a fraudulent one. Banks are switching to EMV cards, smart cards that store data on integrated circuits rather than magnetic stripes as technology progresses, enabling on-card transactions.


### 1.1 MOTIVATION

Fraudsters have improved their techniques over time in order to evade discovery, along with the technologies used to detect fraud. Even though there are many reported research on the use of data mining methodologies for credit card fraud detection, predictive models for credit card fraud detection are still actively used in practice.

### 1.2 OBJECTIVES

Methods for detecting credit card fraud must always be improved. In an effort to better detect credit card fraud, we compare the two cutting-edge data mining techniques known as support vector machines and random forests against the well-known logistic regression.

### 1.3 PROBLEM STATEMENT

Financial loss from fraud is rising dramatically, which has led to a significant increase in credit card scams. Farad LIERTV918070649 causes billions of dollars in losses per year. There isn't enough study to examine the fraud. To find actual credit card fraud in the wild, many machine learning techniques are used.

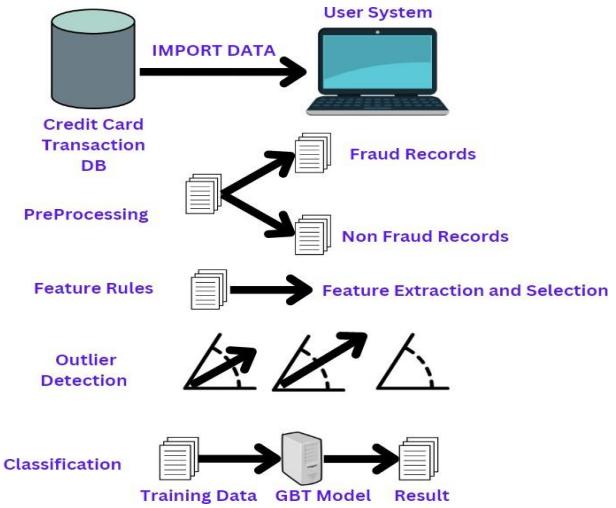
### 1.4 Machine Learning Using Python

A smart and popular programming language is Python. Python supports a wide range of libraries, including pandas, NumPy, SciPy, matplotlib, etc. It supports a variety of packages, including Xlsx, Writer, and X1Rd. It is used to conduct complex science very effectively. There are many useful Python frameworks available. With the use of data, machine learning, a subfield of artificial intelligence, enables computer frameworks to learn new abilities and improve their performance. It is used to conduct research on the creation of computer-based algorithms for data prediction. The process of machine learning begins with the provision of data, after which the computers are taught using a range of algorithms to produce machine learning models

### 1.5 Table Dataset

| SL NO | ATTRIBUTE | DESCRIPTION |
|---|---|---|
| 1 | NAME | HOLDERS NAME ON CREDIT CARD |
| 2 | ACC_NUM | CUSTOMER NUMBER |
| 3 | DATE_TIME | THE TRANSACTION DATE AND TIME |
| 4 | TRX_PER_DAY | DAILY TRANSACTIONS |
| 5 | LOCATION | LOCATION OF THE TRANSACTION |
| 6 | AMT | AMOUNT OF DAILY TRANSACTION |
| 7 | RESULT | VALUE OF THE DECISION |
| 8 | EMAIL AND PH_NO | FALSE OR NOT TO NOTIFY THE PEOPLE OF DECEPTION |

## 2 System Architecture




A Basic System Architecture of Credit Card Fraud Detection Analysis.

### 2.1 Credit Card Fraud Detection

As you can see, there does appear to be a correlation between several of our predictors and the class variable. Despite the large number of factors, there appear to be few significant relationships.

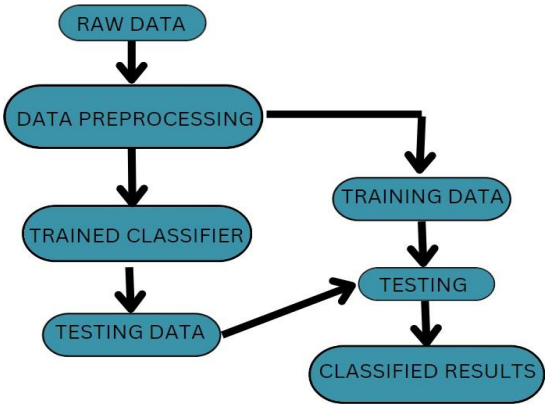
1. Since a PCA was used to prepare the data, our predictors are principal components. 2. Due to the extreme class imbalance, some correlations with respect to our class variable may not be as significant.

### 2.2 About dataset

The files include credit card transactions done by European cardholders in September 2013. We have 492 frauds out of 284,807 transactions in our dataset oftransactions that took place over the course of two days.The dataset is very skewed, with frauds making up 0.172% of all transactions in the positive class. It only has numeric input variables that have undergone PCA transformation. Unfortunately, we are unable to offer the original characteristics and additional context for the data due to confidentiality concerns. The principal components obtained with PCA are features V1, V2,..., V28. The only features that have not been changed with PCA are Time and Amount. The seconds that passed between each transaction and the datasets first transaction are listed in the feature Time.

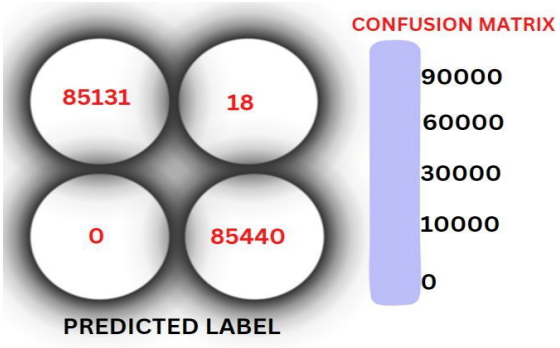
### 2.3 System Framework

It illustrates the steps taken to create the model. The figure illustrates the crucial stages that went into creating the suggested model. Following a series of actions like data processing, data cleansing, and feature extraction, classification is finally carried out.




### 2.4 Results And Discussion

Without a doubt, the Random Forest model outperforms Decision Trees. However, if we look at our dataset, we cansee that there is a significant class imbalance. The number of legitimate (fraud-free) deals exceeds 99, while the number of fraudulent deals is 0.17. If we train our model using a comparable distribution without considering theimbalance problems, it predicts the label with higher significance assigned to actual deals (since there is moreevidence about them) and so acquires additional fragility. There are a variety of viable methods for resolving the class imbalance issue. One of them is overslicing. After oversampling, the accuracy scores and the confusion matrix are calculated.




The Random Forest model performs significantly better than Decision Trees. But our dataset reveals a large class imbalance, as can be seen. There are more than 99 valid (fraud-free) transactions compared to just 0.17 fraudulent transactions. Our model predicts the label with higher importance attributed to actual deals (because there is more data about them) and hence develops additional fragility if we train it using a comparable distribution without taking the imbalance problems into account. There are many effective ways to address the class imbalance problem. One of them is cutting too thin. The accuracy ratings and the confusion matrix are computed following oversampling.

### 2.5 Future enhancements

The accuracy score for our machine model that detects credit card transaction fraud should be 100%, with 100% as the aim. But when we achieve a score of 100% accuracy, we can deduce that our model is being over- fitted with data, giving us the results for which it has already been trained. Therefore, we can say that the precision and confusion matrix values can be enhanced with a wide range for future improvements. The transaction dataset for European credit card holders can also be updated with new algorithms, and the results of these algorithms precision and confusion matrices can be combined to get more accurate numbers. The data collection can also be made better by normalising the extremely skewed numbers and matrices .

## 3 CONCLUSION

The detection of credit card fraud has long been a goal of testing for academics, and it will continue to be an intriguing aspect of testing in the future. By using three different algorithms and training our machine using the transaction information we have, we are launching a fraud detection solution for credit cards. With the aid of the model we developed, the authorities are better able to identify credit card fraud, investigate it further, and determine if it was a fraudulent or valid transaction. These algorithms tell us whether a given transaction has a tendency to be fraud or not; they were chosen using the feature importance, discussion, and experimentation methods as described in the methodology.

## 4 ACKNOWLEDGEMENT

We consider ourselves quite fortunate to have received this support over the course of our project because the success and outcome of this project required a great deal of direction and assistance from many people. We wont forget to thank them for their guidance and support, which are the sole reasons we were able to do what we did. Mr. Puneet Goswami, Head of the Department, Department of Computer Science and Engineering, and Dr. Paramjit S. Jaswal, Vice-Chancellor, SRM University, for giving all the necessary resources for the completion of my seminar. Mrs.Ishwari Singh, who served as our guide, for her insightful advice and help with the study report. Finally, we would want to express our gratitude to everyone.

## 5 REFERENCES

[1] Yvan Lucas, Pierre-Edouard Portier, Lea Laporte , Sylvie Calabretto, Liyun He-Guelon, Frederic Oble and Michael Granitzer , IEEE Explore2019.

[2] Chunzhi, WangYichao, WangZhiwei, YeLingyu , Yuecheng Cai ,The 13th International Conference on Computer Science & Education.

[3] Addisson Salazar, Gonzalo Safont, Luis Vergara, International Conference on Computational Science and Computational Intelligence (CSCI), 2019

[4] Anu Maria Babu , Dr. Anju Pratap , IEEE EXPLORE ,2020

[5] Vinod Jain, Mayank Agarwal , Anuj Kumar , International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) , 2020.

[6] Fahimeh Ghobadi , Mohsen Rohani , ICSPIS ,2016.

[7] SP Maniraj , Aditya Saini, Swarna Deep Sarkar Shadab Ahmed , International Journal of Engineering Research & Technology (IJERT) , 2019..

[8] Adwait A. Rajmane, Piyush S. Mahajan, Akshay D.Kolhe, Sandhya S. Khot , International Journal of Engineering Research & Technology (IJERT) , 2021.

[9] SIMI MJ , International Journal of Engineering Research & Technology (IJERT), 2019.

[10] Arya Chandorkar , JOURNAL-International Journal of Engineering Research & Technology (IJERT), 2022.