IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 3, March 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104190 |

Machine Learning Algorithms for credit card fraudulent Transaction Detection

Dr. M.K. Jayanthi Kannan[1], Varun.A.G[2], Varun.A[3], Venkat Sai[4], Rahul P[5], Sathya Shree[6]

Professor, School of Engineering and Technology, Jain University, Bangalore, India[1]

UG Student, Department of Information Science and Engineering, School of Engineering and Technology, Jain University, Bangalore, India[2, 3, 4, 5, 6]

ABSTRACT: It is important that the goal of data analytics is to discover the hidden patterns and insights that supports informed decision-making. The credit card fraud is a serious problem, and the use of machine learning algorithms to detect fraudulent transactions is becoming increasingly common. One main challenge in detecting credit card fraud is the highly imbalanced nature of publicly available datasets. This means that the number of fraudulent transactions is much lower than the number of legitimate transactions, which can make it difficult for machine learning algorithms to accurately detect the fraud.

However, with the appropriate preprocessing techniques and the use of appropriate machine learning algorithms, it is possible to effectively identify fraudulent transactions. The choice of algorithm depends on the specific characteristics of the dataset, such as the size of the dataset, the number of features, and the degree of class imbalance. In addition to applying supervised machine learning algorithms to detect credit card fraud, it is also important to identify the most important variables that may lead to higher accuracy in fraud detection. This can be done using feature selection techniques, which aims to identify the most relevant features that contribute to the classification task. By removing irrelevant or redundant features, the performance of the classifier can often be improved.

KEYWORDS: Data analytics; Credit card fraud; Supervised machine learning algorithms; Imbalanced datasets; Preprocessing techniques; Feature selection; Performance evaluation

I. INTRODUCTION

Credit card fraud detection is a challenging problem due to the highly imbalanced nature of the dataset. Majority of credit card transactions are legitimate, and only a small percentage are fraudulent. This means that if a model predicts that all transactions are legitimate, it still achieves a high accuracy score, but it fails to detect fraudulent transactions.

To address this issue, one approach is to balance the class distribution in the dataset. One way to do this is to oversample the minority class (fraudulent transactions) to increase the number of training examples. This can be done by randomly duplicating existing examples or generating synthetic examples using techniques such as SMOTE (Synthetic Minority Over-sampling Technique). Another approach is to under sample the majority class (legitimate transactions) by randomly removing examples until the class distribution is balanced. However, this may lead to loss of important information and reduced model performance.

[1] Another approach to handle imbalanced datasets is to use appropriate evaluation metrics such as precision, recall, and F1-score, instead of accuracy. These metrics consider the imbalance in the dataset and provide a better measure of model performance for detecting fraudulent transactions. Additionally, using advanced techniques such as ensemble learning, anomaly detection, and deep learning can also improve the performance of credit card fraud detection models.

II. RELATED WORK

The current state-of-the-art system in credit card fraud detection using machine learning typically involves a combination of supervised and unsupervised learning techniques. The system is typically composed of several components that work together to detect and prevent fraudulent transactions. While machine learning-based credit card fraud detection systems have been successful in identifying fraudulent transactions, there are still some potential disadvantages to these systems. Here are a few of the main disadvantages:

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104190 |

- **Lack of Transparency**: One of the main challenges of machine learning-based systems is that they can be difficult to interpret. It can be challenging to understand why a particular transaction was flagged as fraudulent, which can make it difficult for investigators to follow up on suspicious activity.

- **False Positives:** Machine learning algorithms can sometimes classify legitimate transactions as fraudulent, which can lead to customer frustration and potentially lost revenue for the bank or payment processor. False positives can also create additional work for fraud investigators, who must spend time investigating transactions that turn out to be legitimate.

- **Data Imbalance:** Credit card fraud is a relatively rare event, which can create an imbalance in the dataset used to train the machine learning algorithms. If the dataset is too heavily weighted towards legitimate transactions, the algorithm may not be able to effectively identify fraudulent transactions.

- **Adversarial Attacks:** Machine learning algorithms are vulnerable to adversarial attacks, where fraudsters intentionally manipulate the data to evade detection. For example, a fraudster might attempt to obfuscate their location or change their spending patterns to avoid triggering fraud detection algorithms.

- **Cost:** Building and maintaining a machine learning-based fraud detection system can be expensive, particularly if the system requires a large amount of computing power or extensive training data. These costs may be difficult to justify for smaller financial institutions or merchants.

III. PROPOSED SYSTEM

Using supervised machine learning algorithms to detect credit card fraud and identifying the most important variables is a common approach in this field. Here are some potential benefits of our proposed system:

- **Improved Accuracy:** By using supervised machine learning algorithms, our system learns from labelled data to identify patterns that are indicative of fraudulent behaviour. This results in more accurate fraud detection system than traditional rule-based approaches.

- **Feature Importance:** Identify the most important variables that contribute to the fraud detection helps financial institutions and payment processors understand the underlying drivers of fraud. This information is used to develop more effective fraud prevention strategies.

- **Real-World Dataset:** A real-world dataset makes our model robust and effective in detecting fraud in a variety of contexts. It also helps in identifying the potential biases in the data and ensures that our model is trained on representative samples of transactions.

- **Machine Learning Classifier:** By implementing a machine learning classifier, our system classifies the transactions in real-time, allowing for quick and accurate fraud detection. This helps in preventing fraudulent transactions from being processed, which can save financial institutions and merchants money and prevent customer frustration.

Overall, our proposed system has the potential to improve the accuracy and efficiency of credit card fraud detection using machine learning. However, it is important to address potential challenges such as data imbalances and false positives to ensure that the system is effective in practice.

IV. PROPOSED ALGORITHMS

RandomForest Classifier

```
from sklearn.ensemble import RandomForestClassifier
RF_model=RandomForestClassifier()
RF_model.fit (X_train,y_train)
```

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104190 |

DecisionTree Classifier

```
from sklearn.tree import DecisionTreeClassifier
DTC_model=DecisionTreeClassifier()
DTC_model.fit (X_train,y_train)
```

Naivebayes Classifier

```
from sklearn.naive_bayes import GaussianNB
nb_model=GaussianNB()
nb_model.fit(X_train,y_train)
```

V. SYSTEM ARCHITECTURE

Flowcharts or process maps are graphical representations of a process, showing the inputs, actions, decisions, and outputs involved in the process. They are commonly used to document, analyze, and improve workflows, and are an essential tool for quality control and continuous improvement initiatives. The advantage of using flowcharts in project management is that they provide a clear and concise representation of the entire process. They enable project managers to identify potential bottlenecks, redundancies, and other inefficiencies that can be addressed to improve productivity and reduce costs. Moreover, flowcharts are also helpful in estimating the cost of quality for a particular process. By using the branching logic of the workflow and estimating the expected monetary returns, project managers can evaluate the costs and benefits of different process improvements and prioritize the ones with the highest return on investment. In summary, flowcharts are powerful tools for project management, enabling project managers to visualize and optimize processes, improve quality, and reduce costs.

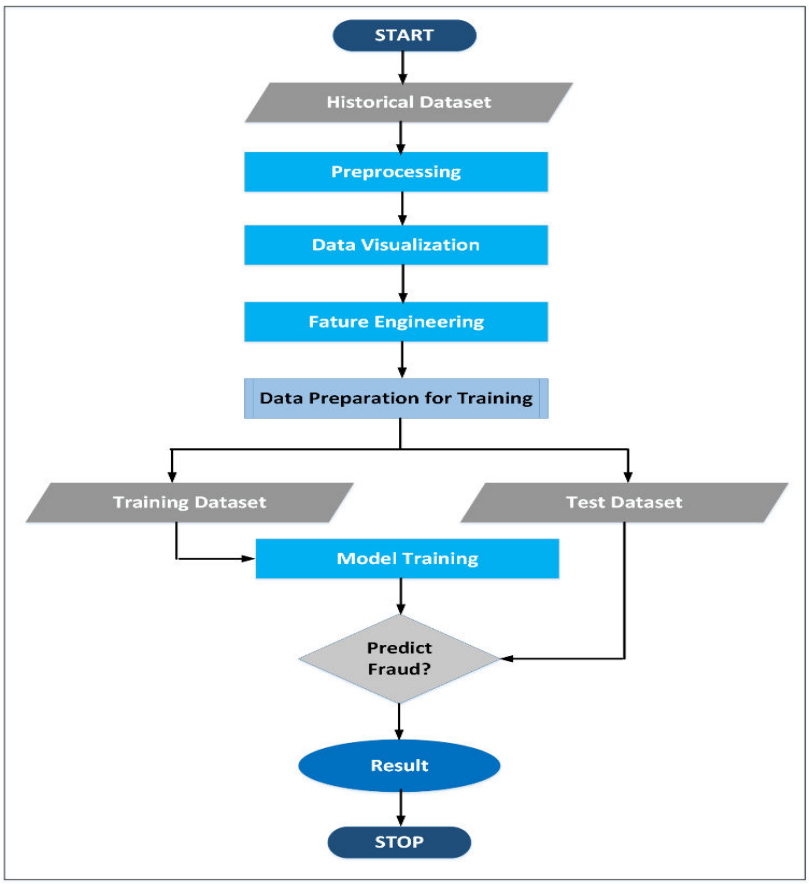

Figure 1. System Architecture

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104190 |

VI. RESULTS

Classification Report of RandomForestClassifier model

	precision	recall	f1-score	support
NotFraud	1.00	1.00	1.00	56618
Fraud	1.00	1.00	1.00	57108
accuracy			1.00	113726
macro avg	1.00	1.00	1.00	113726
weighted avg	1.00	1.00	1.00	113726

Figure 2. Classification Report – RandomForestClassifier

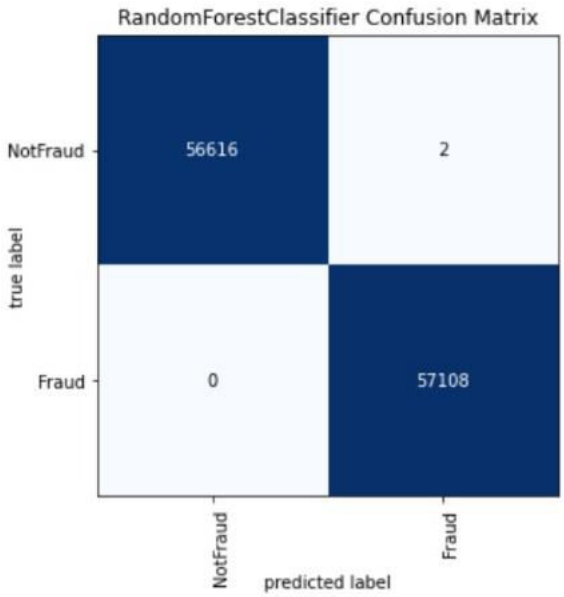

Figure 3. Confusion Matrix – RandomForestClassifier

From figure 2 and 3, it has been concluded that the RandomForest Algorithm achieved 99% accuracy with fi-score 1.00.

Classification Report of DecisionTreeClassifier model

	precision	recall	f1-score	support
NotFraud	1.00	1.00	1.00	56618
Fraud	1.00	1.00	1.00	57108
accuracy			1.00	113726
macro avg	1.00	1.00	1.00	113726
weighted avg	1.00	1.00	1.00	113726

Figure 4. Classification Report – DecisionTreeClassifier

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104190 |

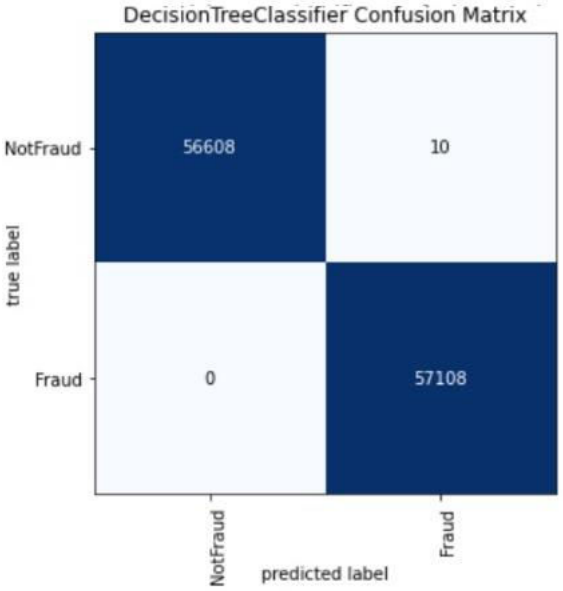

Figure 5. Confusion Matrix –Decision Tree Classifier

From figure 4 and 5, it has been concluded that the Decission Tree Algorithm achieved 99% accuracy with fi-score 1.00.

Classification Report of NaiveBayesClassifier model

	precision	recall	f1-score	support
NotFraud	0.79	0.99	0.88	56618
Fraud	0.99	0.74	0.84	57108
accuracy			0.86	113726
macro avg	0.89	0.86	0.86	113726
weighted avg	0.89	0.86	0.86	113726

Figure 6. Classification Report-Naive Bayes Classifier

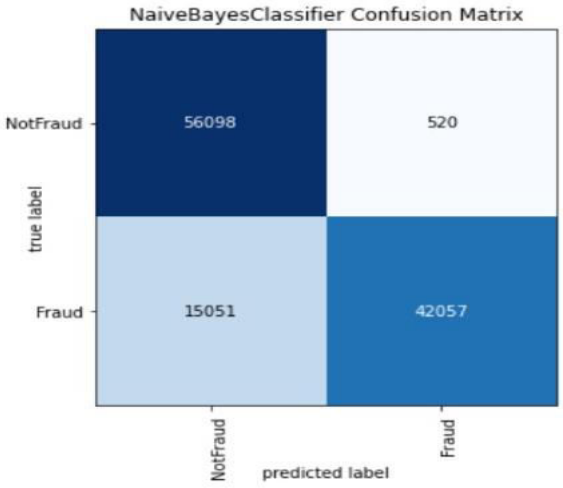

Figure 7. Confusion Matrix-Naive Bayes Classifier

From figure 6 and 7, it has been concluded that the NaiveBayes Algorithm achieved 86% accuracy with fi-score 0.86.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104190 |

VII. CONCLUSION

After comparing three different algorithms, we have concluded that the RandomForest algorithm achieved the result of an accurate value of credit card fraud detection that is 0.9999824126347632. In comparison to the, the RandomForest algorithm, the DecisionTree algorithm and the NaiveBayes algorithm have lower accuracy. The RandomForest algorithm module provides accurate results with larger training data.

REFERENCES

1. S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," Decis. Support Syst., vol. 50, no. 3, pp. 602–613, 2011.
2. K. Chaudhary, J. Yadav, and B. Mallick, "A review of Fraud Detection Techniques: Credit Card," Int. J. Comput. Appl., vol. 45, no. 1, pp. 975–8887, 2012.
3. "Mining of Massive Datasets Second Edition."
4. F. N. Ogwueleka, "Data Mining Application in Credit Card Fraud Detection System," vol. 6, no. 3, pp. 311–322, 2011.
5. H. Nordberg, K. Bhatia, K. Wang, and Z. Wang, "BioPig: a Hadoop-based analytic toolkit for large-scale sequence data,"Bioinformatics, vol. 29, no. 23, pp. 3014– 3019, Dec. 2013.
6. M. Hegazy, A. Madian, and M. Ragaie, "Enhanced Fraud Miner: Credit Card Fraud Detection using Clustering Data Mining Techniques," Egypt. Comput. Sci., no. 03, pp. 72–81, 2016.
7. M. Zareapoor and P. Shamsolmoali, "Application of credit card fraud detection: Based on bagging ensemble classifier," Procedia Comput. Sci., vol. 48, no. C, pp. 679–686, 2015.
8. O. S. Yee, S. Sagadevan, N. Hashimah, and A. Hassain, "Credit Card Fraud Detection Using Machine Learning As Data Mining Technique," vol. 10, no. 1, pp. 23–27.
9. K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, "Credit Card Fraud Detection Using AdaBoost and Majority Voting," IEEE Access, vol. 6, pp. 14277–14284, 2018.
10. N. Mahmoudi and E. Duman, "Detecting credit card fraud by Modified Fisher Discriminant Analysis," Expert Syst. Appl., vol. 42, no. 5, pp. 2510–2516, 2015.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details