IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 9, September 2023

DOI: 10.17148/IJARCCE.2023.12911

Credit Card Fraud Detection

Vijayakrishnan MC[1], Eby Chandra[2], Kumaran M[3]

Student, Department of CSE, Jaya Engineering College, Chennai, India[1]

Assistant Professor, Department of CSE, Jaya Engineering College, Chennai, India[2]

Professor, Department of CSE, Jaya Engineering College, Chennai, India[3]

Abstract: It is vital that credit card companies are able to identify fraudulent credit card transactions so that customers are not charged for items that they did not purchase. Such problems can be tackled with Data Science and its importance, along with Machine Learning, cannot be overstated. This project intends to illustrate the modelling of a data set using machine learning with Credit Card Fraud Detection. The Credit Card Fraud Detection Problem includes modelling past credit card transactions with the data of the ones that turned out to be fraud. This model is then used to recognize whether a new transaction is fraudulent or not. Our objective here is to detect 100% of the fraudulent transactions while minimizing the incorrect fraud classification

Keywords: Credit Card, Card-Present Fraud, Fraud Detection, Card-Not-Present Fraud

I. INTRODUCTION

Credit card generally refers to a card that is assigned to the customer (cardholder), usually allowing them to purchase goods and services within credit limit or withdraw cash in advance. Credit card provides the cardholder an advantage of the time, i.e., it provides time for their customers to repay later in a prescribed time, by carrying it to the next billing cycle. Credit card frauds are easy targets. Without any risks, a significant amount can be withdrawn without the owner's knowledge, in a short period. Fraudsters always try to make every fraudulent transaction legitimate, which makes fraud detection very challenging and difficult task to detect.

II. LITERATURE REVIEW

Multiple Supervised and Semi-Supervised machine learning techniques are used for fraud detection but we aim is to overcome three main challenges with card frauds related dataset i.e., strong class imbalance, the inclusion of labelled and unlabelled samples, and to increase the ability to process a large number of transactions.

Different Supervised machine learning algorithms like Decision Trees, Naive Bayes Classification, Least Squares Regression, Logistic Regression and SVM are used to detect fraudulent transactions in real-time datasets. Two methods under random forests are used to train the behavioural features of normal and abnormal transactions. They are Random-tree-based random forest and CART-based. Even though random forest obtains good results on small set data, there are still some problems in case of imbalanced data. The future work will focus on solving the above mentioned problem. The algorithm of the random forest itself should be improved.

Performance of Logistic Regression, K-Nearest Neighbour, and Naïve Bayes are analysed on highly skewed credit card fraud data where Research is carried out on examining meta-classifiers and meta-learning approaches in handling highly imbalanced credit card fraud data.

Through supervised learning methods can be used there may fail at certain cases of detecting the fraud cases. A model of deep Auto-encoder and restricted Boltzmann machine (RBM) that can construct normal transactions to find anomalies from normal patterns. Not only that a hybrid method is developed with a combination of Ada boost and Majority Voting methods.

III. PROPOSED METHODOLOGY

Card transactions are always unfamiliar when compared to previous transactions made the customer. This unfamiliarity is a very difficult problem in real-world when are called concept drift problems. Concept drift can be said as a variable which changes over time and in unforeseen ways. These variables cause a high imbalance in data. The main aim of our research is to overcome the problem of Concept drift to implement on real-world scenario. Table 1, shows basic features that are captured when any transaction is made.

DOI: 10.17148/IJARCCE.2023.12911

Table 1: Raw features of credit card transactions

Attribute name	Description
Transaction id	Identification number of a transaction
Cardholder id	Unique Identification number given to the cardholder
Amount	Amount transferred or credited in a particular transaction by the customer
Time	Details like time and date, to identify when the transaction was made
Label	To specify whether the transaction is genuine or fraudulent

A. Dataset Description:

The dataset [11] contains transactions made by a cardholder in a duration in 2 days i.e., two days in the month of September 2013. Where there are total 284,807 transactions among which there are 492 i.e., 0.172% transactions are fraudulent transactions. This dataset is highly unbalanced. Since providing transaction details of a customer is considered to issue related to confidentiality, therefore most of the features in the dataset are transformed using principal component analysis (PCA). V1, V2, V3,..., V28 are PCA applied features and rest i.e., ‘time’, ‘amount’ and ‘class’ are non-PCA applied features, as shown in table 2

Table 2: Attributes of European dataset

S. No.	Feature	Description
1.	Time	Time in seconds to specify the elapses between the current transaction and first transaction.
2.	Amount	Transaction amount
3.	Class	0 – Not Fraud 1 - Fraud

B. Methodology

Firstly, we use clustering method to divide the cardholders into different clusters/groups based on their transaction amount, i.e., high, medium and low using range partitioning.

Using Sliding-Window method, we aggregate the transactions into respective groups, i.e., extract some features from window to find cardholder's behavioural patterns. Features like maximum amount, minimum amount of transaction, followed by the average amount in the window and even the time elapsed

Algorithm 1: Algorithm to derive aggregated transaction details and to extract card holder features using sliding window technique.

Input: id of the customer holding a card, a sequence of transactions t and window size w; Output: Aggregated transactions details and features of cardholder genuine or fraud;

```
l: length of T
Genuine= [];
Fraud= [];
For i in range 0 to l-w+1:
        T: [];
       /* sliding window features*/
     For j in range i+w-1:
    /*Add the transaction to window */
    T=T+tjid;
End
```

```
/* features extraction related to amount */
a i1=MAX_AMT(T i );
a i2=MIN_AMT(T i );
a i3=AVG_AMT(T i);
a i4=AMT(Ti );
For j in range i+w-1:
       /* Time elapse */
       x i= Time (t j)-Time (t j-1)
End
X i= (a i1, a i2, a i3,a i4,a i5, );
Y= LABEL(T i );
/* classifying a transaction into fraud or not */
if Yi=0 then
           Genuine =Genuine U X i;
    Else
           Fraud =Fraud U X i;
End
```

Every time a new transaction is fed to the window the old once are removed and step-2 is processed for each group of transactions. (Algorithm for Sliding-Window based method to aggregate are referred from [1]).

Algorithm 2: Algorithm to update the rating score of the classifier to find the accurate the model is.

Input: id of the cardholder and a pervious and a current transaction.

Output: Rating score of the model after every transaction.

T: current transaction with w-1 transaction from window.

C: represents the classifier

Label: true value of the incoming/current transaction.

K: total of transactions processed by model.

```
If the predicted value $\neq$ label and label==0 then,

        For i in range (0, K):
                    if the predicted value $\neq$ label then,
                    rsi= rsi-1;
            Else
                   rsi =rsi+1;
End
```

C. Formula

In our proposed system we use the following formulae to evaluate, accuracy and precision are never good parameters for evaluating a model. But accuracy and precision are always considered as the base parameter to evaluate any model.

The Matthews Correlation Coefficient (MCC) is a machine learning measure which is used to check the balance of the binary (two-class) classifiers. It takes into account all the true and false values that is why it is generally regarded as a balanced measure which can be used even if there are different classes,

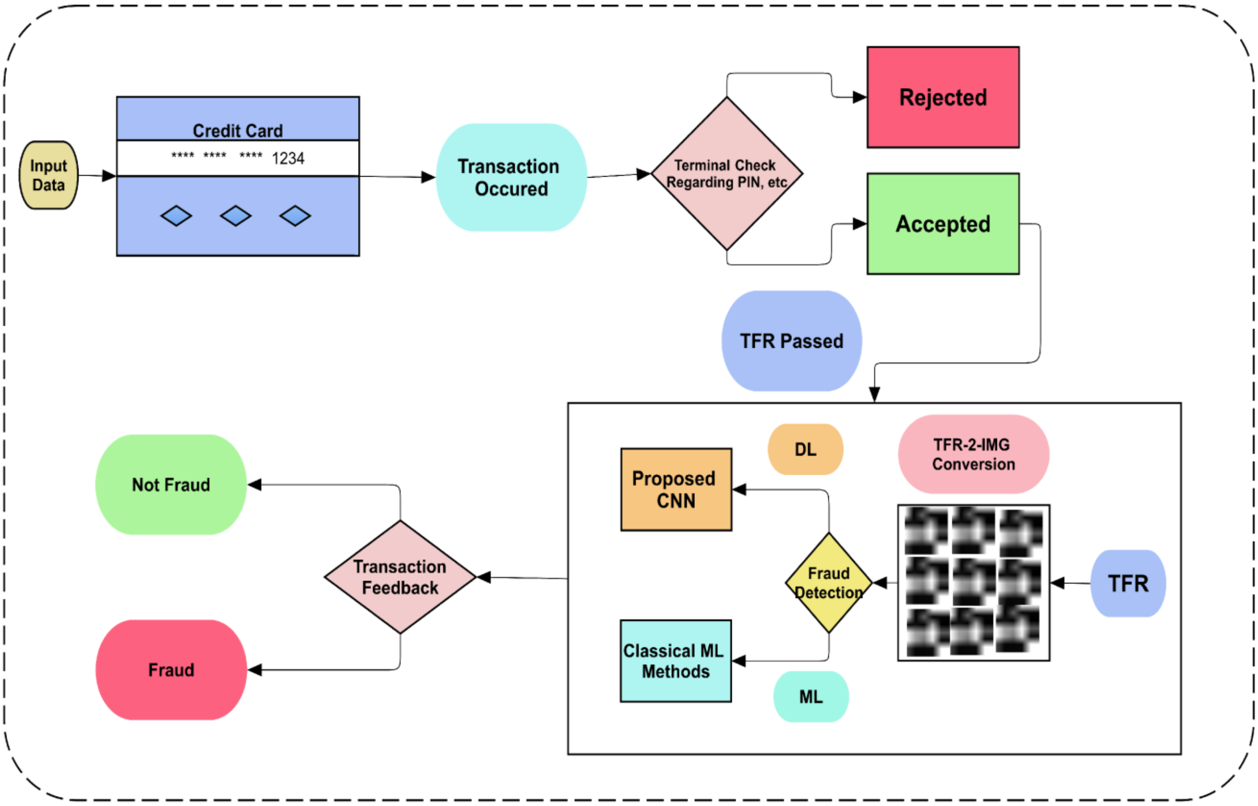
DOI: 10.17148/IJARCCE.2023.12911

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

TP- True Positive
TN- True Negative
FP- False Positive
FN- False Negative



IV. CONCLUSION

In this paper we developed a novel method for fraud detection, where customers are grouped based on their transactions and extract behavioural patterns to develop a profile for every cardholder. Then different classifiers are applied on three different groups later rating scores are generated for every type of classifier.

DOI: 10.17148/IJARCCE.2023.12911

This dynamic change in parameters lead the system to adapt to new cardholder's transaction behaviours timely. Followed by a feedback mechanism to solve the problem of concept drift. We observed that the Matthews Correlation Coefficient was the better parameter to deal with imbalance dataset. MCC was not the only solution. By applying the SMOTE, we tried balancing the dataset, where we found that the classifiers were performing better than before. The other way of handling imbalance dataset is to use one-class classifiers like one-class SVM. We finally observed that Logistic regression, decision tree and random forest are the algorithms that gave better results.

REFERENCES

[1]. Jiang, Changjun et al. "Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism." IEEE Internet of Things Journal 5 (2018): 3637-3647.

[2]. Pumsirirat, A. and Yan, L. (2018). Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine. International Journal of Advanced Computer Science and Applications, 9(1).

[3]. Mohammed, Emad, and Behrouz Far. "Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study." IEEE Annals of the History of Computing, IEEE, 1 July 2018, doi.ieeecomputersociety.org/10.1109/IRI.2018.00025.

[4]. Randhawa, Kuldeep, et al. "Credit Card Fraud Detection Using AdaBoost and Majority Voting." IEEE Access, vol. 6, 2018, pp. 14277–14284., doi:10.1109/access.2018.2806420.

[5]. Roy, Abhimanyu, et al. "Deep Learning Detecting Fraud in Credit Card Transactions." 2018 Systems and Information Engineering Design Symposium (SIEDS), 2018, doi:10.1109/sieds.2018.8374722.

[6]. Xuan, Shiyang, et al. "Random Forest for Credit Card Fraud Detection." 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), 2018, doi:10.1109/icnsc.2018.8361343.

[7]. Awoyemi, John O., et al. "Credit Card Fraud Detection Using Machine Learning Techniques: A Comparative Analysis." 2017 International Conference on Computing Networking and Informatics (ICCNI), 2017, doi:10.1109/iccni.2017.8123782.

[8]. Melo-Acosta, German E., et al. "Fraud Detection in Big Data Using Supervised and Semi-Supervised Learning Techniques." 2017 IEEE Colombian Conference on Communications and Computing (COLCOM), 2017, doi:10.1109/colcomcon.2017.8088206.

[9]. http://www.rbi.org.in/Circular/CreditCard

[10]. https://www.ftc.gov/news-events/press-releases/2019/02/imposter-scams-top-complaints-made-ftc-2018

[11]. https://www.kaggle.com/mlg-ulb/creditcardfraud

[12]. https://www.kaggle.com/uciml/default-of-credit-card-clients-dataset

[13]. https://www.kaggle.com/ntnu-testimon/paysim1/home