International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

Credit Card Fraud Detection Using Artificial Neural Network

S. Charu mathi[1], Mr. S. Arun raj[3], Ms. Sarika jain[3], Dr. S. Geetha[4]

[1]M.Sc – CFIS, Department of Computer Science and Engineering, Dr. M.G.R Educational and Research Institute, Chennai 600 095, Tamilnadu, India

[2,3]Center of Excellence in Digital Forensics, Chennai 600 089, Tamilnadu, India

[4]Head of the Department, Department of Computer Science and Engineering, Dr. M.G.R Educational and Research Institute, Chennai 600 095, Tamilnadu, India

Abstract

The recent advances of e-commerce and e-payment systems have sparked an increase in financial fraud cases such as credit card fraud. It is therefore crucial to implement mechanisms that can detect the credit card fraud. 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 modeling of a data set using machine learning with Credit Card Fraud Detection. The Credit Card Fraud Detection Problem includes modeling 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. The objective here is to detect 100% of the fraudulent transactions while minimizing the incorrect fraud classifications. In the classification process, focused on analyzing and pre-processing data sets as well as the deployment of multiple anomaly detection algorithms such as Local Outlier Factor and Isolation Forest algorithm on the Credit Card Transaction data. So will make use of accuracy and precision to evaluate the performance of the proposed system.

Keywords: Credit card fraud, machine learning.

1. Introduction

Credit card fraud is a form of identity theft that involves an unauthorized taking of another's credit card information for the purpose of charging purchases to the account or removing funds from it. Credit Card Fraud is one of the biggest threats to business establishments today. Credit card frauds are committed in the following ways:

• A demonstration of criminal misdirection (deceive with plan) by utilization of unapproved record or individual data

• Unlawful or unapproved utilization of record for individual addition

• Deception of record data to get products as well as administrations.

2. Literature Survey

Fraud act as the unlawful or criminal deception intended to result in financial or personal benefit. It is a deliberate act that is against the law, rule or policy with an aim to attain

unauthorized financial benefit. Numerous literatures pertaining to anomaly or fraud detection in this domain have been published already and are available for public usage. A comprehensive survey conducted by Clifton Phua and his associates have revealed that techniques employed in this domain include data mining applications, automated fraud detection, adversarial detection. In another paper, Suman, Research Scholar, GJUS&T at Hisar HCE presented techniques like Supervised and Unsupervised Learning for credit card fraud detection. Even though these methods and algorithms fetched an unexpected success in some areas, they failed to provide a permanent and consistent solution to fraud detection. A similar research domain was presented by Wen-Fang YU and Na Wang where they used Outlier mining, Outlier detection mining and Distance sum algorithms to accurately predict fraudulent transaction in an emulation experiment of credit card transaction data set of one certain commercial bank. Outlier mining is a field of data mining which is basically used in monetary and internet fields. It deals with detecting objects that are detached from the main system i.e. the transactions that aren't genuine. They have taken attributes of customer's behaviour and based on the value of those attributes they've calculated that distance between the observed value of that attribute and its predetermined value. Unconventional techniques such as hybrid data mining/complex network classification algorithm is able to perceive illegal instances in an actual card transaction data set, based on network reconstruction algorithm that allows creating representations of the deviation of one instance from a reference group have proved efficient typically on medium sized online transaction. There have also been efforts to progress from a completely new aspect. Attempts have been made to improve the alert-feedback interaction in case of fraudulent transaction. In case of fraudulent transaction, the authorized system would be alerted and feedback would be sent to deny the ongoing transaction. Artificial Genetic Algorithm, one of the approaches that 5 shed new light in this domain, countered fraud from a different direction. It proved accurate in finding out the fraudulent transactions and minimizing the number of false alerts. Even though, it was accompanied by classification problem with variable misclassification costs. Rimpal R. Popat with Jayesh Chaudhary: They made a survey on credit card fraud detection, considering the major areas of credit card fraud detection that are, bank fraud, corporate fraud, Insurance fraud. With these they have focused on the two ways of credit card transactions i) Virtually (card, not present) ii) With Card or physically present. They had focused on the techniques which are Regression, classification, Logistic regression, Support vector machine, Neural network, Artificial Immune system, K-nearest Neighbor, Naïve Bayes, Genetic Algorithm, Data mining, Decision Tree, Fuzzy logic-based system, etc. In which, they have explained six data mining approaches as theoretical background that are classification, clustering, prediction, outlier detection, Regression, and visualization. Then have explained about existing techniques based on statistical and computation which is Artificial Immune system (AIS), Bayesian Belief Network, Neural Network, Logistic Regression, Support Vector Machine, Tree, Self-organizing map, Hybrid Methods, As a result, they had concluded that all the present machine learning techniques mentioned above can provide high accuracy for the detection rate and industries are looking forward to finding new methods to increase their profit and reduce the cost.Machine learning can be a good choice for it. [A Survey on Credit Card Fraud Detection using Machine Learning]. Shiyang Xuan: They made a comparison based on two random forests. Random-tree-based random forest CART-based random forest. They use different random forest algorithms to train the behavior features of normal and abnormal transactions and both of the algorithms are different in their base classifications and their Performance. They applied both of the algorithms on the dataset e-commerce company in China. In which the fraud transaction in the subsets ratio is 1:1 to 10:1. As a result, accuracy from the random-tree based random forest is 91.96% whereas in CART-based random forest is 96.7%. Since the data used is from

the B2C dataset many problems arrived such as unbalanced data. Hence, the algorithm can be improved. [Random Forest for Credit Card Fraud Detection]. Dejan Varmedja: Proposed the various machine learning algorithms and analyzed them concerning to credit card fraud detection methods.

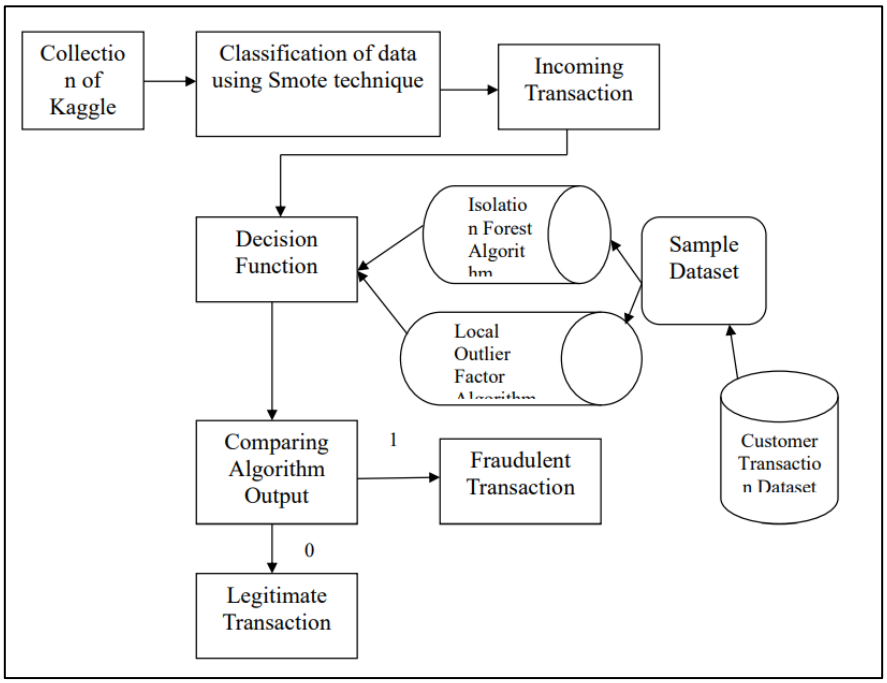
3. Existing System

This existing system uses Artificial Neural Networks and Backpropagation to find fraudulent transactions. This will give the prediction of each transaction and it will detect the fraudulent transactions in real-time. The Credit Card Customer dataset is used that has 31 attributes to ANN model. The first 30 attributes have information related to the customer's Age, Name, Sex, etc. and the last attribute will give the outcome of transactions in either 0 and 1. To customer's confidentiality, the values in our dataset are already transformed using Principal Component Analysis (PCA) technique. These go as an input to the first layer of our multi-layer perceptron. Then the input goes to the first hidden layer. Hidden layer has 15 neurons. In an artificial neural network, given a set of inputs, the activation function of a neuron will define what will the output of a neuron.

4. Proposing System

The Kaggle datasets are trained by using the SMOTE technique. SMOTE technique is used to solve data imbalance problem. Using the smote technique, the data, which is nothing but the transactions are trained. This technique is mainly used to differentiate the fraud transactions from the original transactions done by the card holders. Finally, the smote provide the balance data. 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.

5. Architecture Diagram



International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

5.1 Architecture Diagram

6. List of Phases

There are 6 phases

- Accuracy Analysis
- Error Rate of Classifier
- Transaction Class Distribution
- Amount per transaction by Class
- Time of transaction vs amount by class

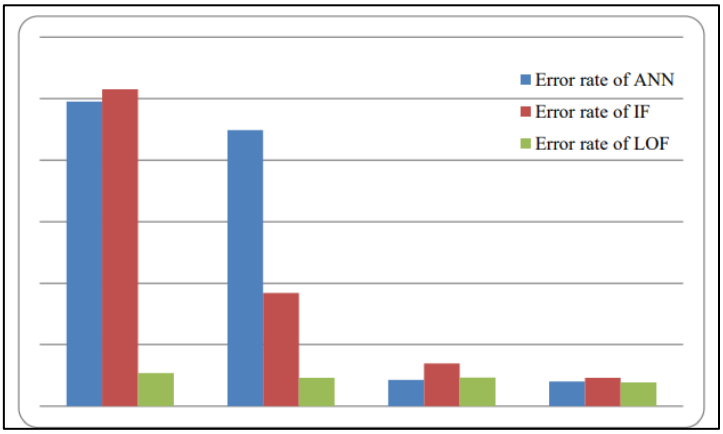
Accuracy Analysis

In this project work we are explain about the credit card spam classification to identify the spam user and non-spam user. For this purpose we are using Isolation Forest and Local Outlier Factor Classifier. In this project we are creating an credit card spam classification system for classify the spam user and non-spam user. Here we are present different reading for all trained dataset which are tested by the classifier i.e. ANN, Isolation Forest and Local Outlier Factor Classifier. Hence shown the different readings and calculation of result:

Train Dataset	Accuracy of ANN	Accuracy of IF	Accuracy of LOF	Error rate of ANN	Error rate of IF	Error rate of LOF
Dataset-50	85	93	98	0.24759	0.25769	0.026923
Dataset-100	82	90	96	0.22456	0.092308	0.023077
Dataset-400	79	88	94	0.02132	0.034615	0.023087

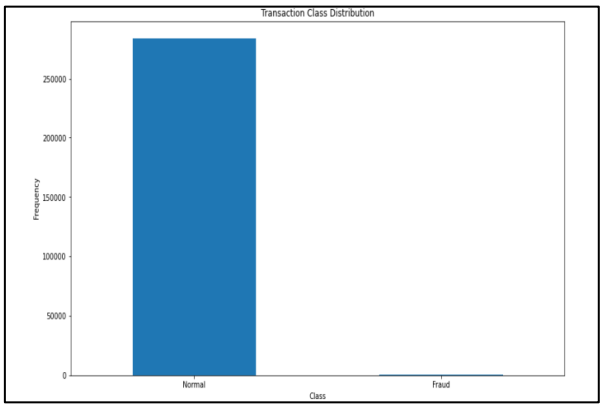
Error Rate of Classifier

In this graph we show the error rate of classifiers for which we calculate the fraction of attribute which are wrongly classified into the total number of attributes. so we can say the LF and ANN provide the high error rate than the LOF Classifier.



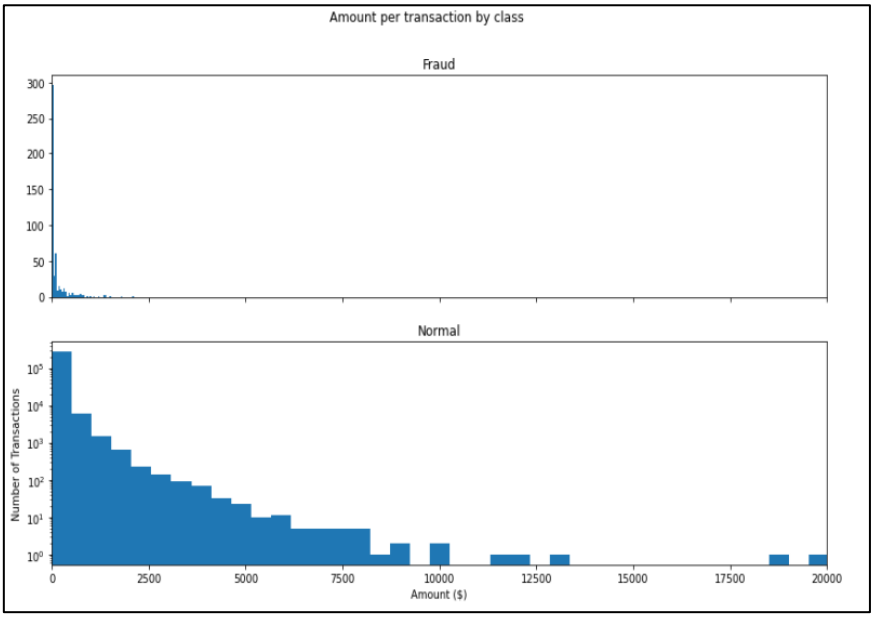
International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

Transaction Class Distribution



This graph shows that the number of fraudulent transactions is much lower than the legitimate ones**.**

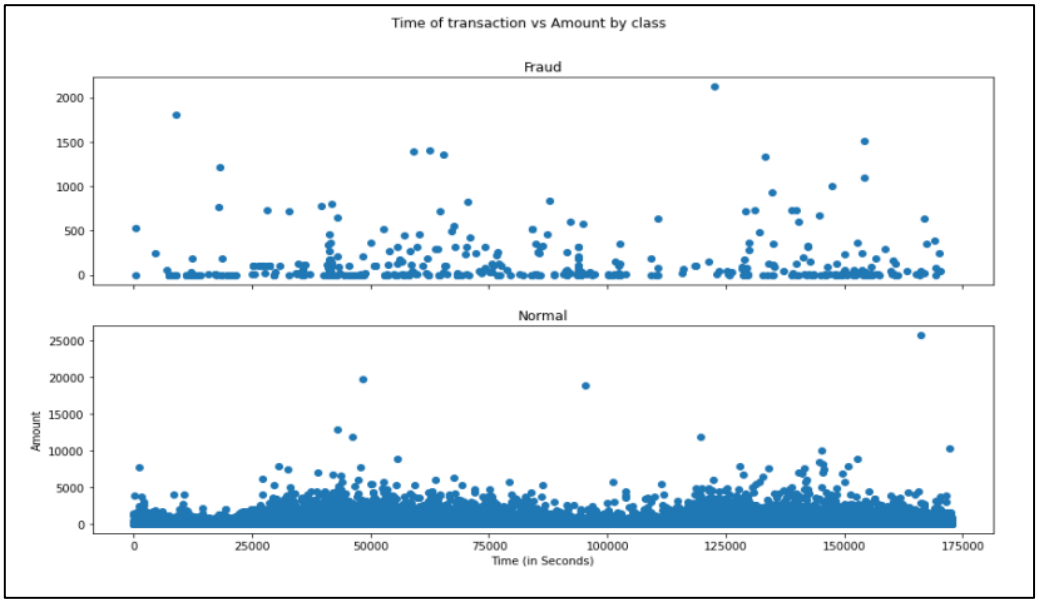
Amount Per Transaction by Class



This graph represents the amount that was transacted. A majority of transactions are relatively small and only a handful of them come close to the maximum transacted amount.

Time of Transaction vs Amount by Class

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795



This graph shows the times at which transactions were done within two days. It can be seen that the least number of transactions were made during night time and highest during the days.

7. Screen Shots

Performance of Classification Algorithms of Local Outlier Factor

```
Local Outlier Factor: 97
Accuracy Score :
0.9965942207085425
Classification Report :
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     28432
           1       0.02      0.02      0.02        49

    accuracy                           1.00     28481
   macro avg       0.51      0.51      0.51     28481
weighted avg       1.00      1.00      1.00     28481
```

Local Outlier Factor detecting 97 errors. Local Outlier Factor has accurate of 99.65%. The error precision & recall for 3 models, the Local Outlier Factor detection rate of just 2 %

Performance of Classification Algorithms of Isolation Forest

```
Isolation Forest: 73
Accuracy Score :
0.9974368877497279
Classification Report :
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     28432
           1       0.26      0.27      0.26        49

    accuracy                           1.00     28481
   macro avg       0.63      0.63      0.63     28481
weighted avg       1.00      1.00      1.00     28481
```

Isolation Forest detected 73 errors versus Local Outlier Factor detecting 97 errors. Isolation Forest has a 99.74% more accurate than LOF of 99.65%. When comparing error precision & recall for 3 models, the Isolation Forest performed much better than the LOF as we can see that the detection of fraud cases is around 27 % versus LOF detection rate of just 2 %. So overall Isolation Forest Method performed much better in determining the fraud cases which is around 30%.

8. Conclusions

Credit card fraud is without a doubt an act of criminal dishonesty. This article has listed out the most common methods of fraud along with their detection methods and reviewed recent findings in this field. This paper has likewise made sense of exhaustively, how AI can be applied to come by improved brings about misrepresentation recognition alongside the calculation, pseudocode, clarification its execution and trial and error results. While the calculation arrives at more than 99.6% exactness, its accuracy stays just at 28% when a 10th of the informational collection is thought about. Nonetheless, when the whole dataset is taken care of into the calculation, the accuracy ascends to 33%. This high level of exactness is not out of the ordinary because of the tremendous awkwardness between the quantity of substantial and number of authentic exchanges. Since the whole dataset comprises of just two days' exchange records, just a small part of information can be made accessible on the off chance that this task was to be utilized on a business scale. Being founded on AI calculations, the program will just expand its effectiveness over the long haul as additional information is placed into it.

References

1. Alrawashdeh K and Purdy C., "Fast Activation Function Approach for Deep Learning Based Online Anomaly Intrusion Detection," 2018 IEEE 4th International Conference on Big Data Security on Cloud (Big Data Security), IEEE International Conference on High Performance and Smart Computing, (HPSC) and IEEE International Conference on Intelligent Data and Security (IDS), Omaha, NE, 2018, pp. 5-13.

2. Amrutha J and Remya Ajai A. S., "Performance analysis of Backpropagation Algorithm of Artificial Neural Networks in Verilog," 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 2018, pp. 1547-1550.

3. Dorronsoro J. R., Ginel F., Sgnchez C. and Cruz C. S., "Neural fraud detection in credit card operations," in IEEE Transactions on Neural Networks, vol. 8, no. 4, pp. 827-834, July 1997.

4. Ghobadi F and Rohani M., "Cost sensitive modeling of credit card fraud using neural network strategy," 2016 2nd International Conference of Signal Processing and Intelligent Systems (ICSPIS), Tehran, 2016, pp. 1-5.

5. Karn S., Sangole S., Gawde A and Joshi J., "Prediction and Classification of Vector-Borne and Communicable Diseases through Artificial Neural Networks," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 2019, pp. 1011-1015.