ISSN: 2321-7782 (Online)
e-ISJN: A4372-3114
Impact Factor: 7.327

Volume 7, Issue 3, March 2019
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Credit Card Fraud Detection

Savitri Deore[1]
Student of Computer Department
Late. G. N. Sapkal College of Engineering
Nashik – India

Prathamesh Burhade[2]
Student of Computer Department
Late. G. N. Sapkal College of Engineering
Nashik – India

Ratnakar Demase[3]
Student of Computer Department
Late. G. N. Sapkal College of Engineering
Nashik – India

Deepali Gaikwad[4]
Student of Computer Department
Late. G. N. Sapkal College of Engineering
Nashik – India

Abstract: Popular payment mode accepted both offline and online is credit card that provides cashless transaction. It is easy, convenient and trendy to make payments and other transactions. Credit card fraud is also growing along with the development in technology. It can also be said that economic fraud is drastically increasing in the global communication improvement. It is being recorded every year that the loss due to these fraudulent acts is billions of dollars. These activities are carried out so elegantly so it is similar to genuine transactions. Hence simple pattern related techniques and other less complex methods are really not going to work. Having an efficient method of fraud detection has become a need for all banks in order to minimize chaos and bring order in place. There are several techniques like Machine learning, Genetic Programming, fuzzy logic, sequence alignment, etc are used for detecting credit card fraudulent transactions. Along with these techniques, outlier detection methods are implemented to optimize the best solution for the fraud detection problem. These approaches are proved to minimize the false alarm rates and increase the fraud detection rate. Any of these methods can be implemented on bank credit card fraud detection system, to detect and prevent the fraudulent transaction.

Key Words: Credit Card Fraud, Classification, Outlier Detection, Precision.

I. Introduction

In day-to-day usage of credit card transactions the procurement of products and services assists online transactions or card swiping procurements. This leads to increase in online transactions using credit and debit cards evolving to a world of effortless expenditure. Frauds involved in the credit card section have caused severe damage to the users and the service provider and is said to be even worse in coming days. Fraudsters observe and adapt to the quick changes in the technology and find clever ways to involve in illegal activities [6]. Frauds caused due to these smart hackers are hazardous and dangerous. A well-educated fraudster can create several identities and conduct credit card transactions without being caught.

Talking in terms of e-commerce transactions the major problem faced due to these fraudulent activities is so similar to legal ones. Hence having an efficient and complex fraud detection system is a must to prevent these fraudulent activities. The challenging section of this problem is to detect frauds in a huge dataset where the legal transactions are more and the fraudulent transactions are bare minimum or close to negligible. There are very few papers on credit card fraud detection methods due to the fact that these methods cannot be tested without a dataset. Hence it’s difficult to prove the robustness or even the probability of success ratio of the methods. As we know that the credit card information is confidential, the bank owners and service providers do not encourage in sharing these data for experiments as well. In this paper we investigate credit card fraud using outlier detection method.

II. Literature Review

1. Abhinav Srivastava et al [1] the author uses the ranges of transaction amount as an attribute in the HMM. The author has suggested method for finding the spending profile of cardholders. It is also discussed how the HMM can identify the fraudulent transactions. The simulation results show the advantages of using HMM and learning the profile of the cardholder plays an important role in analyzing fraudulent cases. The result also shows that 80% of the results are accurate and the system is scalable for large data set as well

2. Divya.Iyer et al [2] the author uses Hidden Markov Model (HMM) to detect credit card transaction frauds. The training set is tuned with the normal behavior of the card holder. So if credit card transaction is rejected by the trained HMM then that transaction is said to be fraudulent. Care is to be taken that valid and genuine transactions are not considered as fraud. The author also compares various methods with the proposed methods to prove that HMM is much preferred than the other methods.

3. K.RamaKalyani et al [3] creates a test data and through which the fraudulent activities are detected. This algorithm is also called as an optimization technique based on genetic and natural selection in high computational problems. The author proposes a method to detect credit card fraud and the results are validated using principles of this algorithm. The purpose of detecting fraud cases is to declare it to the client and the service provider.

4. Renu et al [4] proposed a fraud detection method which involves monitoring the activities of populations to observe and predict undesirable behavior. Undesirable behavior is a set of several habits like intrusion, fraud, delinquency and defaulting. This research speaks on several credit card fraud detection and telecommunication fraud and different techniques which help in resolving the discussed problems.

5. Venkata Ratnam Ganji et al [5] the author uses concept of data stream outlier detection algorithm which is based on anti knearest neighbors for credit card fraud identification. Whereas traditional methods need to scan the database many times to find the fraudulent transaction, which is not suitable for data stream surroundings. This method makes easier to stop fraudulent transaction happens by Lost and stolen card and Credit card validation checks and detects errors in a sequence of numbers which also helps to detect valid and invalid numbers easily.

III. Methodology

3.1 Outlier Detection

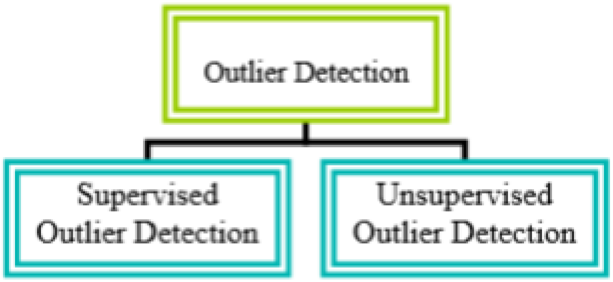


Fig.3.1 Types of Outlier Detection

1. In supervised outlier detection method domain experts model the system to learn and classifier the outlier using training set of data.
2. In unsupervised outlier detection objects are clustered into multiple groups based on features. Those objects are far from any group is labeled as outliers.

The below Fig 3.2 shows the difference between unsupervised and supervised outlier detection mechanism

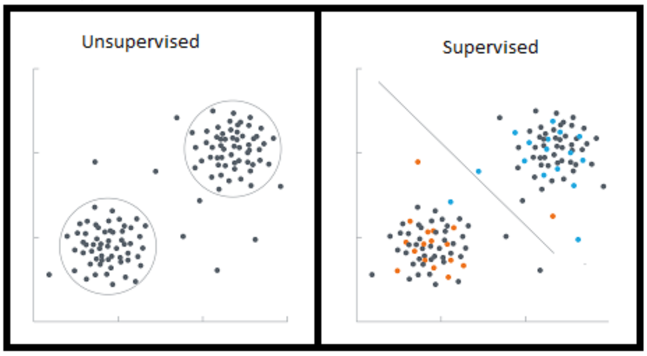


Fig.3.2 Outlier Detection Using Unsupervised and

Supervised methods

In outlier detection method unsupervised learning is preferred to detect the fraud because it can lead to new explanations and representation of the observation data. Another advantage of using unsupervised data, it does not require prior labeling of data or knowledge about fraudulent methods or transactions. So it need not be trained to discriminate between a legal and illegal transaction. It simply follows the normal behavior pattern as an unusual activity or fraudulent. The problem with supervised methods is that the model has to be trained with both fraudulent and non-fraudulent behavior prior to implementing the method in live scenario. Only after the training is performed the system is ready to detect unusual behaviors. The major advantage of using unsupervised method over supervised data is that it need not be trained to discriminate between a legal and illegal transaction.

3.3 Datasets

Throughout the financial sector, machine learning algorithms are being developed to detect fraudulent transactions. In this project, that is exactly what we are going to be doing as well. Using a dataset of nearly 28,500 credit card transactions and multiple unsupervised anomaly detection algorithms, we are going to identify transactions with a high probability of being credit card fraud. In this project, we will build and deploy the following two machine learning algorithms:

- Local Outlier Factor (LOF)
- Isolation Forest Algorithm
- Furthermore, using metrics such as precision, recall, and F1-scores, we will investigate why the classification accuracy for these algorithms can be misleading.
- In addition, we will explore the use of data visualization techniques common in data science, such as parameter histograms and correlation matrices, to gain a better understanding of the underlying distribution of data in our data set.

In the following cells, we will import our dataset from a .csv file as a Pandas DataFrame. Furthermore, we will begin exploring the dataset to gain an understanding of the type, quantity, and distribution of data in our dataset. For this purpose, we will use Pandas' built-in describe feature, as well as parameter histograms and a correlation matrix.

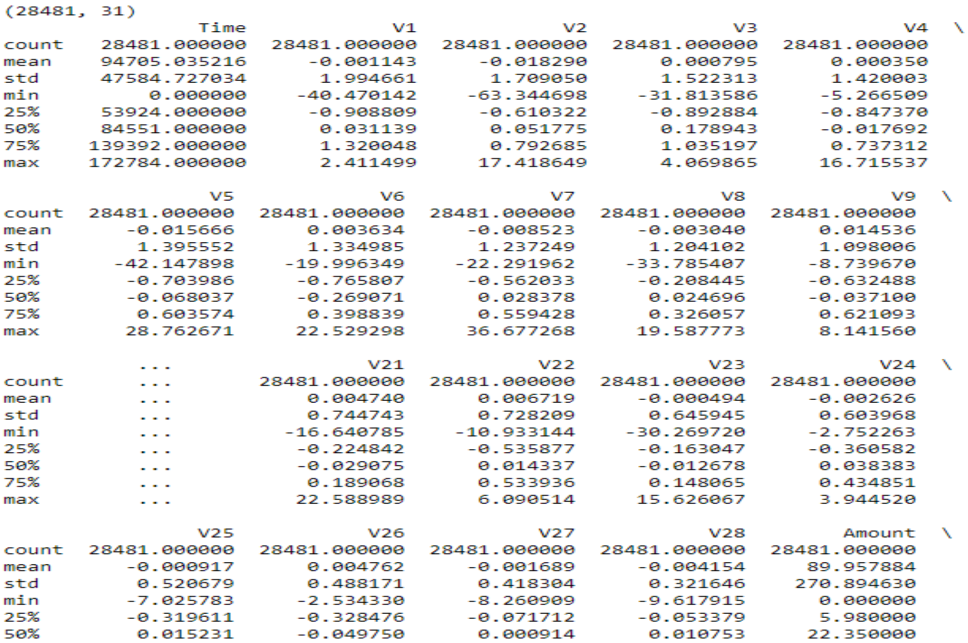
Unsupervised Outlier Detection

Local Outlier Factor (LOF)

The anomaly score of each sample is called Local Outlier Factor. It measures the local deviation of density of a given sample with respect to its neighbors. It is local in that the anomaly score depends on how isolated the object is with respect to the surrounding neighborhood.

Isolation Forest Algorithm

The Isolation Forest 'isolates' observations by randomly selecting a feature and then randomly selecting a split value between the maximum and minimum values of the selected feature. Since recursive partitioning can be represented by a tree structure, the number of splittings required to isolate a sample is equivalent to the path length from the root node to the terminating node. This path length, averaged over a forest of such random trees, is a measure of normality and our decision function. Random partitioning produces noticeably shorter paths for anomalies. Hence, when a forest of random trees collectively produce shorter path lengths for particular samples, they are highly likely to be anomalies.



```
(28481, 31)
                Time            V1            V2            V3            V4  \
count   28481.000000  28481.000000  28481.000000  28481.000000  28481.000000
mean    94705.035216     -0.001143     -0.018290      0.000795      0.000350
std     47584.727034      1.994661      1.709050      1.522313      1.420003
min         0.000000    -40.470142    -63.344698    -31.813586     -5.266509
25%     53924.000000     -0.908809     -0.610322     -0.892884     -0.847370
50%     84551.000000      0.031139      0.051775      0.178943     -0.017692
75%    139392.000000      1.320048      0.792685      1.035197      0.737312
max    172784.000000      2.411499     17.418649      4.069865     16.715537

                 V5            V6            V7            V8            V9  \
count  28481.000000  28481.000000  28481.000000  28481.000000  28481.000000
mean      -0.015666      0.003634     -0.008523     -0.003040      0.014536
std        1.395552      1.334985      1.237249      1.204102      1.098006
min      -42.147898    -19.996349    -22.291962    -33.785407     -8.739670
25%       -0.703986     -0.765807     -0.562033     -0.208445     -0.632488
50%       -0.068037     -0.269071      0.028378      0.024696     -0.037100
75%        0.603574      0.398839      0.559428      0.326057      0.621093
max       28.762671     22.529298     36.677268     19.587773      8.141560

             ...                V21           V22           V23           V24  \
count        ...       28481.000000  28481.000000  28481.000000  28481.000000
mean         ...           0.004740      0.006719     -0.000494     -0.002626
std          ...           0.744743      0.728209      0.645945      0.603968
min          ...         -16.640785    -10.933144    -30.269720     -2.752263
25%          ...          -0.224842     -0.535877     -0.163047     -0.360582
50%          ...          -0.029075      0.014337     -0.012678      0.038383
75%          ...           0.189068      0.533936      0.148065      0.434851
max          ...          22.588989      6.090514     15.626067      3.944520

                V25           V26           V27           V28        Amount  \
count  28481.000000  28481.000000  28481.000000  28481.000000  28481.000000
mean      -0.000917      0.004762     -0.001689     -0.004154     89.957884
std        0.520679      0.488171      0.418304      0.321646    270.894630
min       -7.025783     -2.534330     -8.260909     -9.617915      0.000000
25%       -0.319611     -0.328476     -0.071712     -0.053379      5.980000
50%        0.015231     -0.049750      0.000914      0.010753     22.350000
```

Fig.3.3 Dataset Values

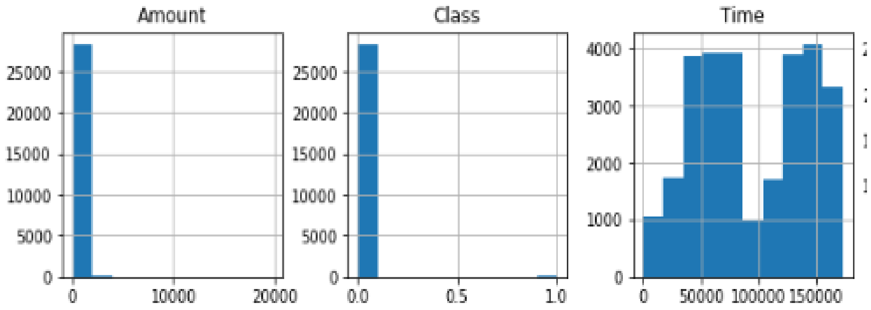


Fig.3.4 Histogram for Amount, Class and Time

IV. CONCLUSION

Credit card scam has become much more extensive. To progress safety measures of the monetary transaction systems in a habitual and effectual way, structure a precise and well organized credit card scam detection system is one of the essential functions for money transactions. By the mean time outlier detection mechanism helps to detect the credit card fraud using less memory and computation requirements. Especially outlier detection works fast and well on online large datasets. But compared with power methods and other known anomaly detection methods.

ACKNOWLEDGEMENT

This research was supported/partially supported by Prof. S.B. Wagh. We are thankful to our colleagues Prof (Dr). N. R. Wankhade head of computer Engineering department and Prof. P. A. Kale who provided expertise that greatly assisted the research, although they may not agree with all of the interpretations provided in this paper.

References

1. Abhinav Srivastava, Amlan Kundu, Shamik Sural, and Arun K. Majumdar" Credit Card Fraud Detection Using Hidden Markov Model" VOL. 5, NO. 1, JANUARY-MARCH 2008 [2]Divya.Iyer,Arti Mohanpurkar,Sneha Janardhan,Dhanashree Rathod,Amruta Sardeshmukh" credit card fraud detection using hidden markov model " 978-14673-0126-8/11/$26.00_c 2011 IEEE
2. K.RamaKalyani, D.UmaDevi" Fraud Detection of Credit Card Payment System by Genetic Algorithm" Volume 3, Issue 7, July-2012
3. Renu, Suman" Analysis on Credit Card Fraud Detection Methods" volume 8 number 1– Feb 2014
4. Venkata Ratnam Ganji," Credit card fraud detection using Anti-k Nearest Neighbor Algorithm",International Journal on Computer Science and Engineering (IJCSE) Vol. 4 ,06 June 2012,(1035-1039)
5. Ekrem Duman, M. Hamdi Ozcelik "Detecting credit card fraud by genetic algorithm and scatter search". Elsevier, Expert Systems with Applications, (2011). 38; (13057–13063).
6. A.J. Graaff A.P. Engelbrecht agraaff "The Artificial Immune System for Fraud Detection in the Telecommunications Environment" 20 November 2014.
7. S. Benson Edwin Raj, A. Annie Portia, "Analysis on Credit Card Fraud Detection Methods", International Conference on Computer, Communication and Electrical Technology – ICCCET2011, 18th & 19th March, 2011
8. Y. Sahin and E. Duman, "Detecting Credit Card Fraud by Decision Trees and Support Vector Machines", International Multiconference of Engineers and computer scientists March, 2011.
9. S. Benson Edwin Raj, A. Annie Portia "Analysis on Credit Card Fraud Detection Methods", IEEE-International Conference on Computer, Communication and Electrical Technology, (2011), pg.152-156.
10. P.Jayant,Vaishali,D.Sharma," Survey on Credit Card Fraud Detection Techniques", International Journal of Engineering Research & Technology (IJERT), Vol. 3 Issue 3, March – 2014,pg.1545-1551
11. Priya Ravindra Shimpi, Prof. Vijayalaxmi Kadroli, "Survey on Credit Card Fraud Detection Techniques", International Journal Of Engineering And Computer Science, Volume 4 Issue 11 Nov 2015, Page No. 15010-15015