International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-10 Issue-5, June 2021

Anomaly Detection Algorithms in Financial Data

Abhisu Jain, Mayank Arora, Anoushka Mehra, Aviva Munshi

Abstract*: The main aim of this project is to understand and apply the separate approach to classify fraudulent transactions in a database using the Isolation forest algorithm and LOF algorithm instead of the generic Random Forest approach. The model will be able to identify transactions with greater accuracy and we will work towards a more optimal solution by comparing both approaches. The problem of detecting credit card fraud involves modelling past credit card purchases with the perception of those that turned out to be fraud. Then, this model is used to determine whether or not a new transaction is fraudulent. The objective of the project here is to identify 100% of the fraudulent transactions while mitigating the incorrect classifications offraud.

***Keywords*: Isolation Forest, Local Outlier, Credit Card, Anomaly Detection**

I. INTRODUCTION

The following are reasons why we need to develop a robust system to detect fraudulent transactions. To make sure that the final product is resilient, the following obstacles have to be addressed. The challenge is to detect fraudulent credit card purchases such that credit card companies' customers are not charged for products they have not purchased.

The main challenges involved in the detection of credit card fraud are:

[1] Enormous data is processed on a regular basis and the design of the model must be rapid enough to respond to the scam in time.

- Imbalanced Data i.e., most of the transactions(99.8percent) are not fraudulent which makes it very hard for detecting the fraudulent ones Data availability as the data is mostly private. Another big problem could be misclassified data, as not every fraudulent transaction is caught and recorded.
- Adaptive techniques used against the model by the scammers.

Manuscript received on May 08, 2021.
Revised Manuscript received on May 12, 2021.
Manuscript published on June 30, 2021.
* Corresponding Author
Abhisu Jain*, Department of Computer Science and Engineering, Vellore Institute of Technology, Vellore (Tamil Nadu), India. Email: jainabhisu007@gmail.com
Mayank Arora, Department of Computer Science and Engineering, Vellore Institute of Technology, Vellore (Tamil Nadu), India. Email: mayankarora220@gmail.com
Anoushka Mehra, Department of Computer Science and Engineering, Vellore Institute of Technology, Vellore (Tamil Nadu), India. Email: anoushka.mehra27@gmail.com
Aviva Munshi, Department of Computer Science and Engineering, Vellore Institute of Technology, Vellore (Tamil Nadu), India. Email: avivamunshi@gmail.com

One of today's greatest risks to business institutions is credit card fraud. Credit card fraud starts either with the theft of a physical card or with essential account-related details, such as the number of a card account or other information that is necessarily open to a merchant during a permissible transaction. Ava strange of metho dsare use dby fraudsters to commit fraud. [2]

The damages resulting from these frauds not only affect financial institutions, but also have a lot of effects on consumers. The identity theft rate remained steady until the mid-2000s, as per the U.S. Federal Trade Commission report, but it rose by 21 percent during2008.

Global card fraud losses rose to US$ 21 billion in 2015, up from about US$ 8 billion in 2010, according to the Nilson Report [1]. The number is anticipated to hit US$ 31 billion by 2020. In order to reduce the losses resulting from these credit card frauds, we need to establish successful strategies.

II. BRIEF ONALGORITHMS

A. Isolation Forest Algorithm

Isolation forest is an unsupervised anomaly detection learning algorithm that operates instead of the most traditional methods of profiling normal points on the concept of isolating anomalies [16]. An anomaly (a.k.a. outlier) is an observation or event in statistics that deviates too much to arouse suspicion that it was cause dbya different means from other events [17]. Anomalies in a large dataset can follow very complex patterns, which in the vast majority of case sare difficult to detect "by eye." [3] This is why the field of identification of anomalies is well adapted to the application of Machine Learning techniques. The most popular anomaly detection techniques are focused on the creation of a profile of what is "usual": anomalies are reported in the dataset as those instances that do not adhere to the normal profile.

B. Local Outlier factor

The Local Outlier Factor (LOF) algorithm is an unsupervised method of detection of anomalies that calculates a given data point's local density deviation with respect to its neighbors.[18] It considers samples that have a slightly lower density than their neighbors as outliers.

Usually, the number of neighbors considered (n neighbors' parameter) is greater than the minimum number of samples to be covered by a cluster, such that other samples may be local outliers compared to this cluster, and 2) lower than the maximum number of sample closures that may theoretically be local outliers. [4]Such knowledge is normal ly not available in practice, and nneighbors=20seems to function well in general.

III. LITERATURESURVEY

In commercial practice, large-scale data-mining techniques will strengthen the state of the art. An important issue, particularly for e-commerce, is scalable techniques to detection task exhibits technical problems, in addition to scalability and performance, which include distorted distribution of training data and non-uniform cost per error, both of which have not been widely studied in the information discovery and data mining community. The suggested methods of integrating multiple learned fraud detectors under a "cost model" are general and demonstrably useful; the empirical results show that through distributed data mining of fraud models we can dramatically reduce losses due to fraud. [6]A neural network-based fraud detection system was trained on a large sample of branded credit card account transactions using data from a credit card issuer and checked on a holdout data set consisting of all account activity over a subsequent two-month period of time. Examples of fraud due to missing cards, stolen cards, application fraud, counterfeit fraud, mail-order fraud and NRI (non-received issue) fraud were trained on the neural network. With significantly less false positives (reduced by a factor of 20) over rule-based fraud detection methods, the network detected considerably more fraud reports (an order of magnitude more). In terms of detection accuracy and early detection of fraud, we discuss the network's performance on this collection of data. [7]Neural Network (NN), rule-induction techniques, fuzzy system, decision trees, Support Vector Machines (SVM), Artificial Immune System (AIS), genetic algorithms, K-Nearest Neighbor algorithms, are the most widely used methods of fraud detection. These methods can be used alone or in collaboration to create classifiers using ensemble or meta-learning techniques. This paper provides a survey of different methods used in the detection of credit card fraud and assesses each methodology on the basis of certain design criteria.[8] A significant application for prediction techniques is the prevention of credit card fraud. [15]The high required diagnostic quality is one major obstacle to the use of neural network training techniques: because only one financial transaction of one thousand is invalid, no predictive performance of lessthan99.9 percent is appropriate. Comprehensive new ideas had to bed eveloped and tested on actual credit card data due to these credit card transaction proportions. The use of credit cards has increased significantly due to the rapid growth of e-commerce technology.[9][10][11] As credit cards are the most common form of payment for both online and daily transactions, there are also growing cases of fraud associated with them. In this paper, a hidden Markov model (HMM) is used to model the sequence of operations in credit card transaction processing and demonstrate how it can be used for fraud detection. Initially, an HMM is educated in the normal actions of a cardholder.[12][13] If an incoming credit card transaction with a reasonably high likelihood is not accepted by a qualified HMM, it is presumed to be fraudulent. This paper addresses automatic detection of credit card fraud using machine learning. [14]Credit card fraud identification is of great significance to financial institutions in an age of digitalization.

IV.EXPERIMENTALRESULTS

A heat-map was generated initially to showcase the strong and weak correlation between different columns of the dataset. The following image shows the heat-map which in turn shows the correlation between different attributes of the dataset.

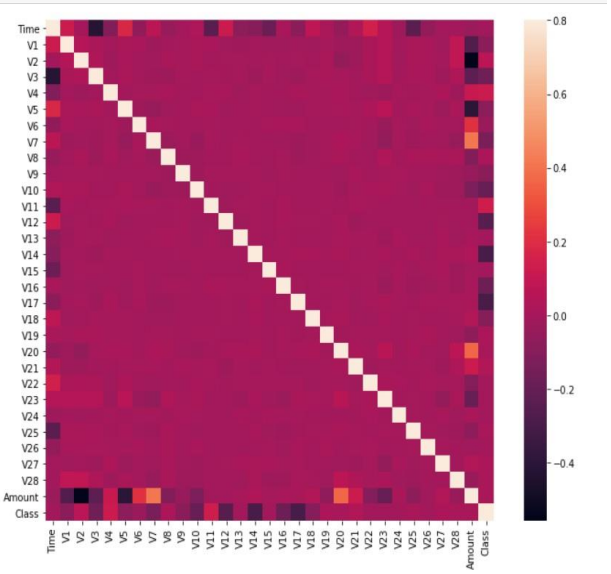

Figure 1: Heat-map depicting correlation

After the correlation was observed both the algorithms were implemented. The following statistics were observed for both algorithms:

Isolation Forest: 71
0.99750711000316

	precision	recall	f1-score	support
0	1.00	1.00	1.00	28432
1	0.28	0.29	0.28	49
accuracy			1.00	28481
macro avg	0.64	0.64	0.64	28481
weighted avg	1.00	1.00	1.00	28481

Local Outlier Factor: 97
0.9965942207085425

	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

Figure 2: Metrics post evaluation

Accuracy of both the models is over 90%, however, Isolation forest algorithm is much better and has a greater precision as compared to LOF algorithm

V. CONCLUSION

For the process of anomaly detection, Isolation forest algorithm does a better job in finding anomalies as compared to LOF algorithm and has a greater precision score. Hence, Isolation forest algorithm is more reliable as compared to LOF algorithm despite the fact that both the algorithms have accuracies greater than 90 %.

REFERENCES

1. Chan,P.K.,Fan,W.,Prodromidis,A.L.,&Stolfo,S.J.(1999). Distributed data mining in credit card fraud detection. IEEE Intelligent Systems and Their Applications, 14(6)W.-K.Chen,*Linear Networks and Systems*
2. Ghosh, S., &Reilly, D. L. (1994, January). Credit card fraud detection with a neural-network. In System Sciences, 1994. Proceedings of the Twenty-Seventh Hawaii International Conference on (Vol. 3, pp. 621-630).IEEE
3. Raj, S. B. E., & Portia, A. A. (2011, March). Analysis on creditcardfrauddetectionmethods.In2011International Conference on Computer, Communication and Electrical Technology (ICCCET) (pp. 152-156).IEEE
4. Brause,R.,Langsdorf,T.,&Hepp,M.(1999,November). Neural data mining for credit card fraud detection. In Proceedings 11th International Conference on Tools with Artificial Intelligence (pp.103-106)
5. Srivastava, A., Kundu, A., Sural, S., & Majumdar, A. (2008). Credit card fraud detection using hidden Markov model. IEEE Transactions on dependable and secure computing,5(1)
6. Maes, S., Tuyls, K., Vanschoenwinkel, B., &Manderick, B. (2002, January). Credit card fraud detection using Bayesian and neural networks. In Proceedings of the 1st international naiso congress on neuro fuzzy technologies
7. Aleskerov, E., Freisleben, B., &Rao, B. (1997, March). Cardwatch: A neural network based database mining system for credit card fraud detection. In Proceedings of the IEEE/IAFE 1997 computational intelligence for financial engineering (CIFEr) (pp. 220-226).IEEE
8. Quah, J. T., &Sriganesh, M. (2008). Real-time credit card fraud detection using computational intelligence. Expert systems with applications,35(4)
9. Ogwueleka, F. N. (2011). Data mining application in credit card fraud detection system. Journal ofEngineering Science and Technology,6(3)
10. Ogwueleka, F. N. (2011). Data mining application in credit card fraud detection system. Journal ofEngineering Science andTechnology
11. Syeda, M., Zhang, Y. Q., & Pan, Y. (2002, May). Parallel granular neural networks for fast credit card fraud detection. In 2002 IEEE World Congress on Computational Intelligence
12. Dal Pozzolo, A., Caelen, O., Le Borgne, Y. A., Waterschoot,S.,&Bontempi,G.(2014).Learnedlessons in credit card fraud detection from a practitioner perspective. Expert systems withapplications
13. Stolfo, S., Fan, D. W., Lee, W., Prodromidis, A., & Chan, P. (1997, July). Credit card fraud detection using meta-learning: Issues and initialresults.
14. Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P. E., He-Guelton, L., &Caelen, O. (2018). Sequence classification for credit-card fraud detection. Expert Systems withApplications
15. Carneiro, N., Figueira, G., & Costa, M. (2017). A data mining basedsystem for credit-card fraud detection in e-tail. Decision SupportSystems
16. Zhu, Kevin. "Phone usage pattern as credit card fraud detection trigger.

AUTHORS PROFILE

Abhisu Jain, UG Final year Student at Vellore Institute of Technology, actively pursuing Computer Science and Engineering. Core Interests: Researching actively in the application of Machine Learning Algorithms in financial world. Email ID: jainabhisu007@gmail.com

Mayank Arora, UG Final year Studentat Vellore Institute of Technology, actively pursuing Computer Science and Engineering. Core Interests: Machine Learning and Artificial Intelligence.
Email ID: mayankarora2205@gmail.com

Anoushka Mehra, UG third year Student at Vellore Institute of Technology, actively pursuing Computer Science and Engineering. Core Interests: Machine Learning, Augmented and Virtual Reality.
Email Id: anoushka.mehra27@gmail.com

Aviva Munshi, UG third year Student at Vellore Institute of Technology, actively pursuing Computer Science and Engineering. Core Interests: Machine Learning, Natural level processing.
Email Id: avivamunshi@gmail.com