A Comparative Study on Credit Card Fraud Detection

Hardik Jethava1
Department of Master of Computer Applications
School of CS & IT
Jain (Deemed-to-be-University)
Bangalore, India

Feon Jaison[2]
Associate Professor
Department of Master of Computer Application
School of CS & IT
Jain (Deemed-to-be-University)
Bangalore, India

Abstract **-** *Financial fraud is an ever threat to the financial sector, with far-reaching consequences. Data mining was crucial in detecting credit card fraud in internet transactions. Credit card fraud detecting, which is a data mining challenge, is a problem due to two fundamental causes., becomes difficult: First, legitimate and fraudulent behaviour profiles are continually changing, and second, credit card fraud data sources are highly skewed. The method used to sample the dataset, the parameters chosen, and the identification method(s) used all have an impact on the accuracy of fraud detection in credit card transactions. This study examines the performance of multiple classification algorithms on heavily skewed credit card fraud data. A total of 284,807 credit card transactions were collected from European consumers in this study. A hybrid method of being under and oversampling is utilised on the skewed statistics. The three techniques are applied to the raw and normalised data. Python is utilised to complete the task. An accuracy, sensitivity, specificity, and precision of the procedures are all examined.*

Keywords – Dataset, Credit card, Random Forest Algorithm, Logistic Algorithm.

I. INTRODUCTION

Financial fraud is an ever-increasing issue in the financial industry, business organisations, and government, by far consequences. Fraud is defined as the intentional use of deception to earn financial gain. The growing reliance on web - based technologies has resulted in an increase in credit card transactions. Credit card fraud is on the rise as credit cards be the most used way of paying for both physical and digital purchases. There are two kinds of credit card fraud: internal card fraud and external card fraud. Internal card fraud occurs when consumers and banks cooperate to commit fraud using a false identification, whereas exterior card fraud occurs when a lost credit card is used to gain cash in questionable methods. Exterior card fraud is the most common type of credit card fraud, has attracted a great deal of investigation. It takes a long time and is inefficient to detect fraudulent transactions using manual system detection methods, as result of the arrival of big data, manual procedures have become outdated. And from the other hand, financial companies have concentrated their efforts on current computational methods to combat credit card fraud. One significant way for detecting credit fraud is the use of data mining techniques. There are two types of techniques for identifying fraudulent transactions: legitimate (genuine) and illegal transactions. Detecting credit card fraud involves looking at a card's purchasing behaviour

II. LITERATURE SURVEY

Sonal Mehndiratta [1] developed "Credit Card Fraud Detection Techniques" In this paper, this research looks at a variety of malware detection systems based on established of parameters. Predictive analysis approaches can be used to detect fraud by collecting historical data. Hidden Markov Designs, Genetic Algorithms, Convolutional Neural Networks, Naive Bayes, and KNN classifiers are among the methods employed. In this work, fraud prediction is primarily accomplished through two phases: feature extraction and classification, In the upcoming, it has been planned to use a hybrid technique for detecting credit card fraud.

Kuldeep Randhwa Et.al [2] established "Credit card fraud detection using AdaBoost and majority voting" developed a machine learning-based method for detecting credit card fraud. Models that are popular were utilised at first, However, hybrid systems such as AdaBoost and qualified majority techniques arose. A publicly available data set was utilised to assess the model's effectiveness, While the scam was being investigated, other set of data from the banking institution was utilised. After then, the noise was applied to data set, allowing the algorithms' toughness to be assessed. The methods were focused on theoretical findings that show the majority of voting systems have good accuracy rates when it came to finding credit card fraud. For further evaluation of the hybrid models, noise of between 10% and 30% has been placed in the sample dataset. For 30 percent increased noise, several voting techniques received a good score of 0.942. As a result, it was established that the voting procedure performed well in the face of noise.

Krishana Modi.Et.al [3] developed "Review on Fraud Detection Methods in Credit Card Transaction" The study looked into numerous techniques for detecting fraudulent transactions and compared them. To detect fraudulent activity, any of these tactics, or a mixture of them, can be utilised. It may be feasible to improve the model's accuracy by introducing new features. For detecting fraud behaviours, Data gathering is used by banks and other financial institutions in a different format. Based on previous behaviours, any of these approaches can be used to establish a client's normal usage behaviour. As a result, a survey of different detection techniques that have been given across time is undertaken here

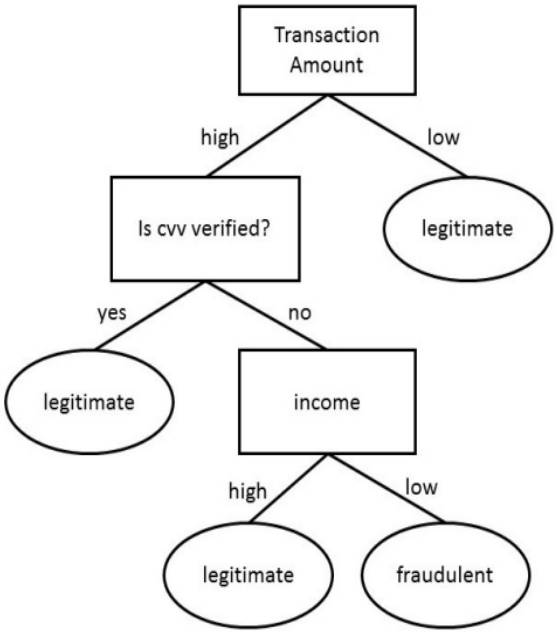

Figure 1 Example of Decision Tree

Figure 1 depicts the difference between a genuine and a fraudulent transaction.

John O. Awoyemi [4] developed "Credit card fraud detection using machine learning techniques: A comparative analysis" designed a study in which the efficiency of a variety of algorithms was evaluated as they were applied to credit card fraud data that was substantially distorted. the set of data credit card transactions was created using the 284,807 transactions of European cardholders as a source. On the skewed statistics, a hybrid strategy of being under and oversampling is utilised. On raw and post data, Python employs three different strategies. The performance of these techniques is assessed using parameters such as precision, sensitivity, accuracy, balanced classification rate, and so on. In terms of performance, the results suggest that k-NN beats naive Markov and logistic regression approaches.

B. Comparison Of different methods

Methods	Speed of detection	accuracy	cost
HMM	Fast	Low	High expensive
FDS	Very low	Very high	High expensive
AIS	Very fast	Good	Inexpensive
FNN	Very fast	Good	Expensive
NN	Fast	Medium	Expensive
DT	Fast	Medium	Expensive
BN	Very Fast	High	Expensive
KNN	Good	Medium	Expensive
SVM	Low	Medium	Expensive
SOM	Fast	Medium	Expensive
BP	low	Low	Expensive
GA	Good	Medium	Inexpensive

Table-1 Comparison of different methods

Table-1 depicts a comparison of multiple methods, taking into account detection speed, accuracy, and cost.

III. CONCLUSION & FUTURE SCOPE

We began by importing a csv set of data. pre-processed it, and then explored and described the information. In addition, a histogram was used to examine for unusual parameters that were essential to our class. Isolation forest and local outlier factor are two techniques used to find anomalies. There is a dataset. We grasped the significance of data comprehension and precision. In terms of precision, number of errors, precision, f1 and recall scores, we see that Isolation Forest outperforms Local Outlier Factor. In the future, we can use Neural Network classifier to train our system to better its accuracy [5]. We began by importing a

csv set of data, pre-processing it, and then exploring and describing the data. In addition, a histogram plot is used to look for anomalous characteristics. We created a correlation matrix to determine which parameters were most essential to our class. Isolation forest and local outlier factor are two techniques used to find anomalies. We understood the value of precision and data understanding in the dataset. Detection of fraud is a difficult topic that demands careful planning before implementing machine learning approaches. Nonetheless, it is an excellent application of machine learning and data science because it assures that the client's money is safe and secure. The system will be applied in the future, with neural networks being used to train the machine for better efficiency. Having a data collection containing non-anonymized features would add to the intrigue, as presenting selected features would allow one to determine which individual qualities are most important for identifying fraudulent transactions.

REFERENCES

[1] Anuruddha Thennakoon; Chee Bhagyani; Sasitha Premadasa; Shalitha Mihiranga; Nuwan Kuruwitaarachchi 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence);10-11 Jan. 2019
[2] John O. Awoyemi; Adebayo O. Adetunmbi; Samuel A. Oluwadare; 2017 International Conference on Computing Networking and Informatics (ICCNI) 29-31 Oct. 2017.
[3] Dejan Varmedja; Mirjana Karanovic; Srdjan Sladojevic; Marko Arsenovic; Andras Anderla; 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH) 20-22 March 2019
[4] Pranali Shenvi; Neel Samant; Shubham Kumar; Vaishali Kulkarni; 2019 IEEE 5th International Conference for Convergence in Technology (I2CT) 29-31 March 2019 [5] Deepti Dighe; Sneha Patil; Shrikant Kokate; 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)16- 18 Aug. 2018
[6] Krishna Modi; Reshma Dayma; 2017 International Conference on Intelligent Computing and Control (I2C2); 23-24 June 2017
[7] S P Maniraj;Aditya Saini;Shadab Ahmed ; International Journal of Engineering and Technical Research 08(09); September 2019;vol 08;page no. 110-115.
[8] S. Abinayaa, H. Sangeetha, R. A. Karthikeyan, K. Saran Sriram, D. Piyush; International Journal of Engineering and Advanced Technology (IJEAT) ;4, April, 2020; vol 09; page no. 1199-1201.
[9] K.Ratna Sree Valli , P.Jyothi , G.Varun Sai , R.Rohith Sai Subash; Quest Journals Journal of Research in Humanities and Social Science; 2 June 2019;Volume 8; page no: 04-11 .
[10] Lakshmi S V; Selvani Deepthi Kavila; International Journal of Applied Engineering Research ISSN; 04 November 2018; Volume 13, pp. 16819-16824