Journal of
Web Development and Web Designing

Volume-5, Issue-2 (May-August, 2020)

www.matjournals.com

Detection of Credit Card Attackers using Machine Learning

N. Snehalatha[1], Aishwarya Damodaran[2], Amulya H N[2*], Apoorva N[2], Aruna M B[2]
[1]Assistant Professor, Department of Computer Science, JSS Academy of Technical Education, Bangalore, Karnataka, India
[2]UG Student, Department of Computer Science, JSS Academy of Technical Education, Bangalore, Karnataka, India

*Corresponding author: amulyanarayan.159@gmail.com

ABSTRACT

E-commerce or online shopping websites use electronic payment. This mode of payment has brought about a revolution in business and almost all industrial sectors. Any small-scale industries or start-ups take to online advertisements and business as a means to popularize their ideas with minimum investment. It has successfully flourished as one of the best methods of business. No business can be perfect online because everything comes with a prize, and in this case, it is mostly all the problems that arise with respect to online transactions which are mostly done by credit cards. This project presents a survey of various techniques used in credit card fraud detection mechanisms and evaluates each methodology based on certain design criteria.

Keywords-- Artificial intelligence, credit card, E-commerce, fraud

INTRODUCTION

Fraud has been increasing drastically with progression of technology. As the concept of online shopping as well as virtual money transactions has gained popularity in the recent days, people use electronic cards on a daily basis to pay school fees, shopping bills, electricity bills, etc. which in turn creates an opportunity for hackers to indulge in theft and perform fraudulent transactions. Thus, financial fraud has become an issue for consumers, businesses as well as financial industries. Credit card fraud is mainly done for financial gains by an unauthorized person without the knowledge of the card holder. The owner of the card is unaware until he/she receives a statement of transactions from the bank. According to a report Credit card fraud was ranked number one kind of Identity theft fraud - accounting for 35.4 per cent of all identity theft fraud in 2018 [1]. Fraud can be avoided either by prevention or by detection. Prevention is a method which tries to avoid fraudulent transactions while Detection is a method used when Prevention fails. Detection acts as an alerting system as well as helps to identify fraudulent transactions. In general machine learning techniques as well as data mining algorithms can be used to detect fraud. A large number of credit card fraud detection systems also use artificial intelligence techniques and pattern matching to identify fraud. There are many types of Credit card fraud that can take place. Card-not-present and Card-present frauds are the two main types of frauds. One of the main solutions to the above problem is to use Machine learning algorithms as they can work on large amount datasets which is humanly not possible. Machine learning techniques can be divided into two main categories: supervised learning and unsupervised learning. Fraud detection can be done in any of the above stated ways and the type used can be decided based on the dataset available. Supervised learning technique will be used for this purpose, but to try and do so we have got to perform classification first. In our paper, we explain the method of detecting fraud by Genetic Algorithm to form a well-organized and dependable payment system. Here, we make use of machine learning techniques to create clusters of training data and identify the spending profile of the cardholder. Our system is unaware of the number of purchased goods as well as the types of goods that are purchased in a particular transaction but it concentrates mainly on the amount of goods purchased and makes use of this information for further processing. Here, data is stored in the form of clusters of different amounts of transactions depending on the transaction amount which will be categorized in ranges of high, medium or low values.

LITERATURE REVIEW

Many approaches are proposed in previous studies to detect fraud from MasterCard transactions like: supervised, unsupervised, or hybrid approaches. Fraud patterns have changed with time and has introduced us to newer varieties of fraud. The rest of this section describes machine

learning models and algorithms for detecting fraudulent transactions. Implementing efficient machine learning models has been the most problem that we face during fraud detection. They mention that the Lack of real-life data is a big issue due to data sensitivity and privacy issues [2, 3]. Papers, studied skewed distribution of data as there is fewer fraudulent data when compared to non-fraudulent data in the transaction datasets [3, 4]. Paper, looks into the issue where data mining techniques takes more time to execute when big data comes into picture [3]. Overlapping of data is also another problem that might arise during preparation of transaction data. Papers, looks into the issue where some legitimate transactions look exactly like fraudulent ones also vice versa [4, 5]. Paper, informs us about the importance of having a great understanding of performance measure [6]. Many different types of models are implemented for fraud detections. In those models, different types of algorithms also come into picture. Adapting the system to frauds that have been newly introduced can also be a problem [7]. Another type of model called Risk-Based Ensemble (RBE) can handle data with issues and also give results.

To handle noise in transaction dataset, Naive Bayes algorithm is used [8]. Credit card data is of skewed distribution type which can also be referred to as class imbalance [9]. Chee et al. used hybrid methods and twelve standard models using AdaBoost for better accuracy [10]. Paper, shows the outstanding performance for sensitivity and specificity using KNN algorithm [11]. The paper talks about commonly used supervised techniques as well as provides a thorough evaluation of supervised learning techniques [12]. The proposed system in paper, is designed in such a way that it can handle class imbalance, formation of unlabelled and labelled data as well as processing of huge datasets [13]. The proposed system was able to overcome all the challenges.

EXISTING SYSTEM

With the current day scenario and improvements in technology online transactions are popular and hence leading to a rapid growth in the amount of frauds related to virtual money and one click transactions. To prevent these kinds of illegal or unwanted activity it is important that every bank or online vendor implements a Fraud Detection system. Illegal transactions are termed as fraud because they happen without the permission or sometimes even knowledge of the card owner/holder. There are mainly the following types of detection approaches: misuse detection and anomaly detection. In misuse detection, the model learns or trains on usual and unusual transactions, it tries to spot or detect the unusual or suspicious transactions. In anomaly or difference detection system normal or usual transactions are used for learning or in other words training. anchez ´ et al. described the procedure to detect frauds from transactional databases using fuzzy association rule mining in extracting knowledge. This method is seen to be very efficient and drastically improves the time required for execution. Panigrahi et al. proposed the new approach using rule-based filtering [14].

Disadvantages of the Existing System

- The main disadvantage of the existing system is the detection occurs only after a written complaint is registered.
- In the existing system physical inconvenience exists.
- The time period required to detect the fraud is high and, therefore, the cardholder might incur a lot of loss by then.
- There is no particular security system in the existing model so a hacker can easily access others card [14].

METHODOLOGY

One of the ways of detecting credit card fraud is using Machine Learning. This technique can be used to identify various suspicious activities. Records of all the previous transactions are maintained in a database and if any unusual transaction is carried out, i.e., which varies from a regular pattern, then such transactions are recorded. The user is informed about such unusual transactions by sending an alert to their mobile phones along with the details of the alarming transaction. The Credit Card Fraud Detection System running at the bank server or on the merchant site records the particulars of the items purchased and these details can be obtained to warn the customer. The implementation techniques of machine learning in order to detect fraud transaction through credit cards, it creates clusters of training set and identify the spending profile of cardholder [15].

Architecture Diagram:

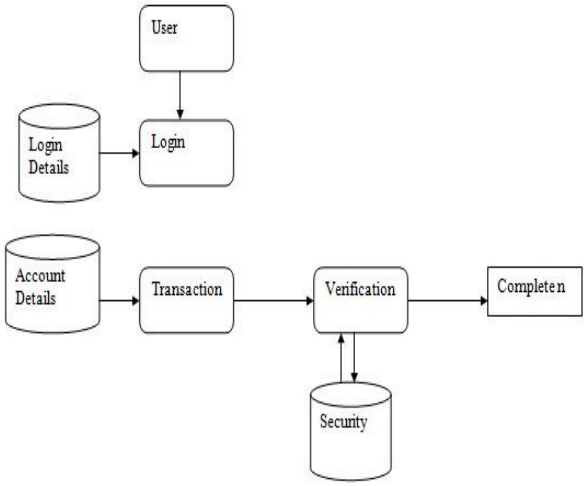

Figure 1: *Architecture diagram.*

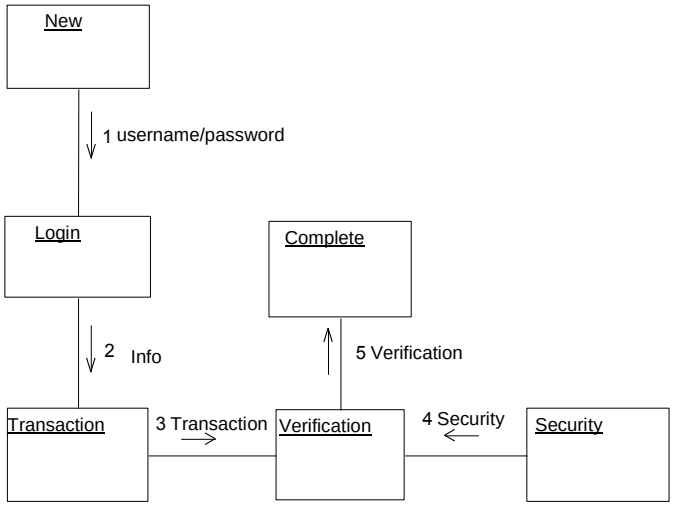

Figure 2: *Methodology.*

MODULES AND DESCRIPTION

Modules

- New card
- Login
- Security information
- Transaction
- Verification

Module Description

New Card

In this module, the customer gives their information to enroll a new card. The information is all about their contact details. The users will have to create their username and password that is unique, and use it to login.

Login

Login Form is used to obtain the username and password from the site visitors as shown in Fig. 1. After validating the username and password, the site visitors will be given access to some of the additional resources on the website. These additional resources will be configured separately based on the user's requirement.

Security Information

The Security information module will get the details from the user stores it in the database. If the card is lost then the Security information module form arises. It contains a set of question which the user has to answer them correctly to proceed to the transaction section. It is responsible to provide privacy and security to user's information and also provides informational self-determination and these are addressed fully by this invention which provides persons and entities a confidential means to users where they can secure, search, process, and exchange personal and/or confidential information [13].

Transaction

A communication device is available between the client and the merchant with the end goal of pre-approval of the exchanges. The Mastercard proprietor starts a Mastercard exchange utilizing their charge card number and putting away in that, an extraordinary snippet of data that describes a particular exchange to be made by an approved client of the Visa for any further exchanges. This data is gotten as "organize information" in the database just if a right close to home ID code (PIC) is utilized while imparting. The "arrange information" will serve to later approve and verify that particular exchange. The Mastercard proprietor or other approved clients can at exactly that point continue with that particular exchange with the Mastercard. Since the exchange is preauthorized, the seller does not have to see or transmit a PIC [16].

Verification

In Fig. 2, the classified data of the starting party held by the outsider is utilized to check the data shared between a starting gathering and the confirmation looking for party. This is accomplished for each exchange. During confirmation, the card number is checked and on the off chance that the card number is right, at that point the significant procedure will be executed. In the event that the number is not right, at that point a mail will be sent to the client saying their card has been blocked and they would not have the option to do promote exchanges.

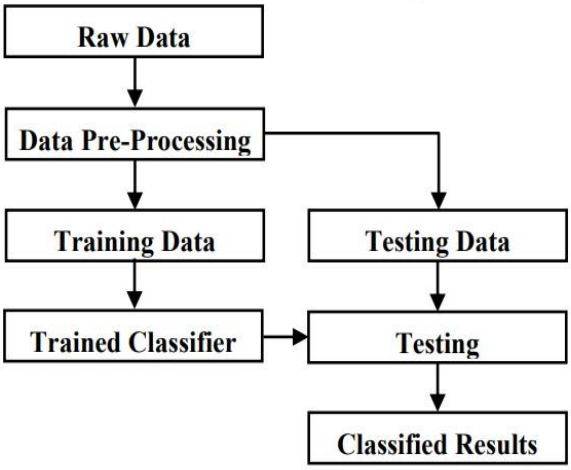

Figure 3: *Workflow diagram.*

RESULTS

In Fig. 3, the detection process constitutes of four steps. These steps are mentioned below:

- All transaction records are given as input which includes confidential information like credit card number, time of transaction and location.
- In the second step, Card usage frequency count, Carid location, Current bank balance and also average daily spending pattern is computed.
- Using Genetic algorithm, we calculate critical fraud, monitorable and ordinary frauds.
- In the last step we find fraudulent transactions by applying detection mining technique.

The proposed system can overcome this problem, in an effective manner. In this way, the algorithm detects fraud and generates efficient result.

The results largely show than compared to the unsampled dataset, the dataset unsampled with ADASYN had much worse performance. Additionally, the unsampled dataset using alternate class weighting to favour the minority class had superior performance than both the unsampled and

ADASYN sampled datasets. This in fact contradicts current research which showed that ADASYN lead to increased overall performance in fraud detection versus other techniques [15].

There are three suspected reasons that explain the discrepancy.

- The generated data was not representative of true fraudulent transactions. This is possible in three parts: ADASYN failed to create synthetic samples which captured the underlying characteristic of fraudulent transactions, or that fraudulent transactions, at least in the used dataset, do not have strong enough differentiation when compared to legitimate transactions. Or, the class imbalance was simply too great to mitigate using ADASYN. However, the positive results obtained using the un-sampled and un weighted methods suggest that the classes are in fact separable. Therefore, it is suspected that ADASYN did not manage to create new fraudulent samples that were representative of the data. Whether or not this is due completely to the extreme class imbalance or due to the nature of the data itself is unknown.
- The specific implementations of the SVM, RF, and MLP algorithms used were not able to generalize to highly imbalanced test and validation data, after training with equally balanced data generated by synthetic sampling. This is supported by the high FP rate when ADASYN is used.
- It is possible that the Python2.7 package used for the ADASYN algorithm, "Imbalance-Learn" contained errors in the implementation that led to the poor results observed. The package is flagged as "experimental" and "use at your own risk".

Comparing the performance of the different classifiers used, it is clear that linear SVM produces the best performance over RF and MLP. In such case it was the un sampled training data with class reweighting that obtained optimal classification. MLP trained with un-weighted and un sampled data came in second but with a FN rate three times high compared to SVM.

RF came closely in third using un sampled training data with class reweighting (Fig. 4).

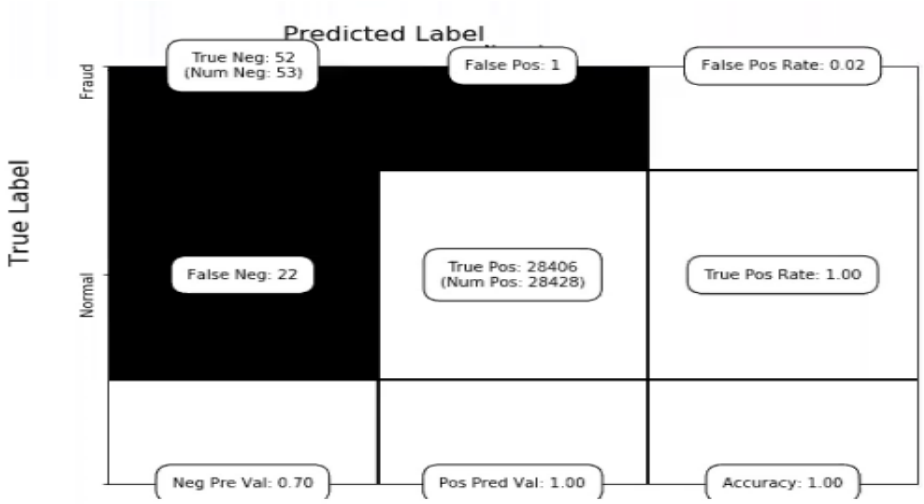

Figure 4: *Confusion matrix.*

The accuracy of the model is found to be 99.91%.

CONCLUSION

The credit card fraud detection system presents an effective system to identify frauds involved in hacking credit cards. It is becoming an important topic of research as the number of cyber security attacks are increasing at an alarming rate. In this paper, it is clearly shown that various algorithms used can detect the fraud efficiently and provide accurate security. This method provides accurate results in finding out the fraudulent transactions and minimizing the number of false alerts. It is clear that when compared to conventional methods for dealing with class imbalance such as cost-based method results in much worse performance when compared to the working of present credit card fraud detection system.

FUTURE WORKS

The inferences obtained here is not in a generalized form. It cannot be directly used in the global fraud detection problem as we have considered a sample data set. For further enhancement we can make use of some effective

algorithm for classification problem with variable misclassification costs.

REFERENCES

1. D. S. Sisodia, N. K. Reddy, S. Bhandari (21–22 September, 2017), "Performance evaluation of class balancing techniques for credit card fraud detection", *IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, Chennai, India.
2. M Rafało (2017), "Real-time fraud detection in credit card transactions", *Data Science Warsaw*.
3. David Robertson (2016), "Investments & amp; Acquisitions–September 2016 top card issuers in Asia-Pacific card fraud losses reach $21.84 billion", Nilson Report, Number1096.
4. Z. Zojaji, R. E. Atana, A. H. Mondesi (2016), "A survey of credit card fraud detection techniques: Data and technique-oriented perspective", *Crypt. and Sec*. pp. 1–26.
5. M. Zareapoor, Seeja K. R and M. Afshar Alam (2012), "Analysis on credit card fraud detection techniques: based on certain design criteria", *Inter. J. of Comp. App*., Volume 52, Issue 3, pp. 35–42.
6. J. West, M. Bhattacharya (5–7 June, 2016), "An investigation on experimental issues in financial fraud mining", *IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)*, Hefei, China.
7. M. F. Zeager, A. Sridhar, N. Fogal, et al. (28 April, 2017), "Adversarial learning in credit card fraud detection", *Systems and Information Engineering Design Symposium (SIEDS)*, Charlottesville, VA, USA.
8. S. Akila, U. S. Reddy (23–24 November, 2017), "Risk based bagged ensemble (RBE) for credit card fraud detection", *International Conference on Inventive Computing and Informatics (ICICI)*, Coimbatore, India.
9. A. Roy, J. Sun, R. Mahoney, L. Alonzi, et al. (27 April, 2018), "Deep learning detection fraud in credit card transactions", *Systems and Information Engineering Design Symposium (SIEDS)*, Charlottesville, VA, USA.
10. K. Randhawa, C. K. Loo, M. Seera, et al. (2018), "Credit card fraud detection using AdaBoost and majority voting", *IEEE Access*, Volume 6, pp. 14277–14284, DOI: 10.1109/ACCESS.2018.2806420.
11. J. O. Awoyemi, A. O. Adetunmbi, S. A. Oluwadare (29–31 October, 2017), "Credit card fraud detection using machine learning techniques: A comparative analysis", *International Conference on Computing Networking and Informatics (ICCNI)*, Lagos, Nigeria.
12. R. Choudhary, H. K. Gianey (14–15 December, 2017), "Comprehensive review on supervised machine learning algorithms", *International Conference on Machine Learning and Data Science (MLDS)*, Noida, India.
13. G. E. Melo-Acosta, F. Duitama-Munoz and J. D. Arias-Londono (16–18 August, 2017), "Fraud detection in big data using supervised and semi-supervised learning techniques", *IEEE Colombian Conference on Communications and Computing (COLCOM)*, Cartagena, Colombia.
14. Suraj Patil, Varsha Nemade, Piyush Kumar Soni (2018), "Predictive modelling for credit card fraud detection using data analytics", *Proced. Comp. Sci.*, Volume 132, pp. 385–395, DOI: 10.1016/j.procs.2018.05.199.
15. Mishra Jaba, Panda Soumyashree, Mishra Ashis (2013), "A novel approach for credit card fraud detection targeting the Indian market", *Inter. J. of Comp. Sci*., Volume 10, Issue 3, pp. 172–1 79.
16. Pappa Kani. P, Mahalakshmi. T, Kavi Priya. M (2016), "Credit card fraud detection system", *Inter. J. of Adv. Res. in Bio. Eng. Sci. and Tech.*, Volume 2, Special Issue 15, pp. 293–296.