International Journal of Advance Research in Science and Engineering
Volume No. 11, Issue No. 04, April 2022
www.ijarse.com
IJARSE
ISSN 2319 - 8354

CREDIT CARD FRAUD DETECTION SURVEY

Akanksha Mirgane[1], Rajvi Landge[2], Aiman Shaikh[3], Mrs.Alfiya Shahbad[4]

[1,2,3,4]Computer Department, Trinity Academy of Engineering, Pune, India

Abstract:

After demonetization and virus breakdown it was observed that most of the deals were continued in a cashless manner around the world. This led to increased payment fraud, especially in credit cards. Therefore, we performed this ANN technique to draw attention to credit card fraud. We analyzed that every fourth person in the country is a victim of credit card fraud, which was reported in a cyber-crime investigation cell. We evaluated victim and fraudulent spending habits, threshold value, a location that causes credit card fraud. We found that most frauds were done during online shopping as the customer was supposed to add card details in payment apps. The model was trained to keep track of actual owner and fraudulent owner spending habits, the irregular transaction was found then ANN blocks the card for further transactions. To detect this fraud, we are explaining methods like Clustering, Artificial Neural Networks. Credit card fraud has increased in past years. Particularly, in online payment; therefore customers should exercise extreme caution to reduce fraud.

Keywords: Credit card fraud, Clustering, Ann Algorithm, Machine Learning, Random Forest.

I. INTRODUCTION

After demonetization and the virus spread out, the country was exercising internet transaction facilities. According to the survey on digital payments in 2020, it was seen that 30 percent of people practice cashless payment through debit or credit cards. As the country was looking forward to digitalization, payment activity supported the country with internet payment. As many people exercised their payment through credit cards this led to increasing fraudulent use of credit cards.

To reduce credit card fraud ANN technique was applied. As the model is used to describe the events that depend on an internal factor that is unknown to the user. Fraud can be done on the virtual card, the fraudulent user needs to know the information about the credit card such as CVV n, Secure code, Credit card number. Therefore, a secure payment gateway is needed to identify the user and to verify that the user is legal or an attacker. The most useful and appropriate technique used for fraud detection is a neural network based on clustering.

The motivation of the report is that as technology advances so do the risk associated with these transactions. The world is so used to online transactions because it is easy to practice. So, it is essential that we need to be very cautious about the increased fraud activities.

The scope of the report is to prevent the fraudulent user from using the actual user's credit card credentials. As the fake user and actual user spending patterns will be different that will be spotted by this model. If the fake user attempted to misuse the credentials the model will alert the actual user and the analyst will block the card immediately.

The objectives of fraud detection are to provide security to customers at the time of transaction. Create a database that contains all relevant information about customers. [1]

The ANN algorithm will be the most effective method to counter fraud transactions through the internet. If it catches an irregular pattern it will automatically block the card and send a notification to the actual user

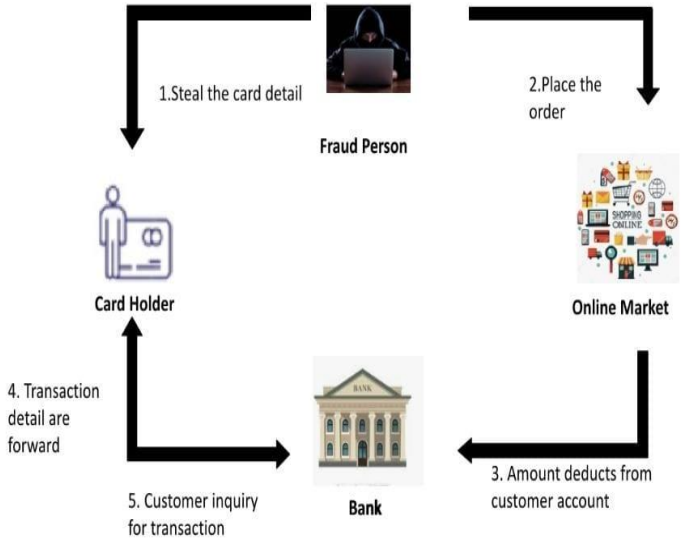

Figure 1: Transaction Flow

II. LITERATURE SURVEY

The literature survey gives the information that using various techniques the fraud has been minimal. Which is stopping the fraudulent user to misuse the cash of the actual user. Though it was very difficult to spot the difference between actual users and fraudulent users. But now the detection system is detecting fraud quickly and not considering genuine users the fraudulent user.

Credit card fraud detection has received significant consideration from researchers around the world. Several techniques have been developed to detect fraud using a credit card which is based on bayesian networks, neural networks, data mining, clustering techniques, genetic algorithms, decision trees, etc.

1.Comparative Analysis of Back-propagation Neural Network and K-Means Clustering Algorithm in Fraud Detection Online credit Card transactions was proposed by Abdulsalami, B.A., Kolawole, A.A., Ogunrinde, M.A., Lawal, M., Azeez, R.A.and Afolabi, A.Z in the year 2019 to have a halt in online transactions. This paper presents that BPNN and K-means algorithms are used in fraud detection in an online credit card transaction. Where BPNN is of great efficiency and least lenient to raise alarm compared to K-means.[2]

2. According to Credit Card Fraud Discovery: A Survey, the researcher Ritika Wadhwa has done a study on five vital fraud detection techniques i.e. Credit Card Fraud detection through Decision Trees, Genetic Algorithms, Neural Networks, Hidden Markov Model and Support Vector Machines. Where a decision tree is a data representation that comprises a root node, branches, and leaf nodes that depend upon Depth-First or Breadth-First Approach. This reduces the complexity of the data sets by partitioning the unknown data sets. The genetic

algorithm goes on checking Credit Card frequency, location, overdraft bank balance, and daily spending to give the best and most optimized solution to the problem. Neural Network runs as the human mind works. An artificial neural network with some training detects the fraud by having the information of the actual user i.e .income, frequency of practicing card, and location of purchase. The hidden Markov model detects fraud when fraud has occurred. Support Vector Machine is to detect a hyperplane that depends on several features.[3]

III. FRAUD DETECTION TECHNIQUES:

1. ANN TECHNIQUE :

An artificial Neural Network is a process that mimics the way that the human brain operates. ANN is rarely used for predictive modeling.

It combines the thinking power of the human brain with the computational power of the machine. The previous year's data is fed into the network and then based upon that data it recognizes a new incoming transaction to be a fraud or a genuine one

A. Existing Model

In existing system methods such as neural network based on clustering is used to spot credit card fraud. In this system, the fraud was detected after the fraud was taken place or if someone reports it. By using discriminate analysis and regression analysis the fraud was detected considering the credit card rate and credit card transaction of the actual user.[4]

And so the cardholder faced a lot of trouble before the investigation finished. And also as all the transaction is maintained in a log, we need to maintain huge data. And also nowadays a lot of online purchases are made so we don't know the person who is using the card online, we just capture the IP address for verification purposes. So they need help from cybercrime to investigate the fraud. To avoid the entire above disadvantage we propose the system to detect fraud in the best and easy way.

B. Proposed Model

The proposed system overcomes the above-mentioned problem efficiently. It aims to analyze the number of fraud transactions that are present in the table. This system analyzes different techniques of fraud transactions which provides the user a secure environment while payment activities. In this system, the process of distance sum into credit card fraud detection is used.

The details of items purchased in Individual transactions are usually not known to any Fraud Detection System(FDS) running at the bank that issues credit cards to the cardholders. Hence, we feel that ANN is an ideal choice for addressing this problem. Another important advantage of the ANN-based approach is a drastic reduction in the number of False Positives transactions identified as malicious by an FDS although they are genuine. An FDS runs at a credit card issuing bank.

C. FDS Working

Fraud Detection System does the following:

a. Each incoming transaction is submitted to the FDS for verification.

b. FDS receives the card details and the value of the purchase to verify, whether the transaction is genuine or not.
c. The types of goods that are bought in that transaction are not known to the FDS.
d. It tries to find any anomaly in the transaction based on the spending profile of the cardholder, shipping address, billing address, etc.
e. If the FDS confirms the transaction to be fraud, it raises an alarm, and the issuing bank declines the transaction.[3]

D.Algorithm

a. Collects all previous data and ask some question.
b. Data that is collected gets processed.
c. Data gets into SVM (Support Vector Machine) and Ann model to get classified.
d. After classification, the data is then tested.
e. If yes, then the transaction is done by a genuine user.
f. If no, then the transaction is done by a fraudulent user.[5]

E.Execution

a. The customer places an order on a merchant's or e-commerce website after this customer's information is sent from EBS to acquiring bank.
b. The acquiring bank then sends the information for authentication.
c. Now Ann's model works in the following order :

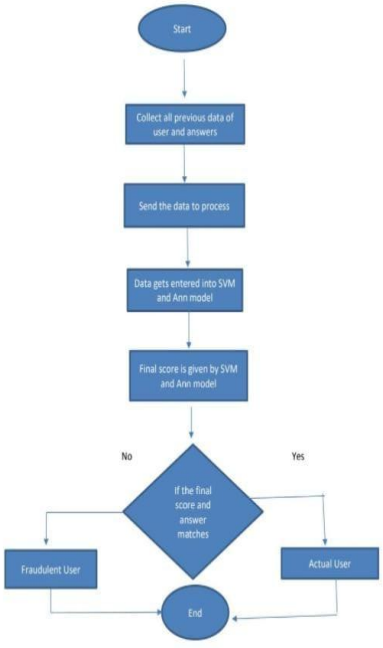

Figure 2: Flowchart

2. RANDOM FOREST :

The Random Forest algorithm is one of the widely used supervised learning algorithms. This can be used for both regression and classification purposes. But, this algorithm is mainly used for classification problems. Generally, a forest is made up of trees and similarly, the Random Forest algorithm creates the

decision trees on the sample data and gets the prediction from each of the sample data. Then Random Forest algorithm is an ensemble method. This algorithm is better than the single decision trees because it reduces the over-fitting by averaging the result.

1. Login :

The Ann model asks the customer to fill in his bank account information like CVV(Credit Verification Value) number, OTP(One Time Password) number, credit card account number, and password, and then the purchased amount is displayed.

2. Security Information :

In this phase the customer has to go through a question and answer round, these answers are compared with the customer's previous answer. If the comparison matches then the customer is allowed to move further for the transaction.

3. Spending Profile :

This module keeps a record of the customer's average amount per transaction, average daily spending habits and times of using a credit card, and so on.

4. Verification :

It verifies whether the entered data is correct or incorrect to identify the actual and fraudulent user.

5. Transaction :

This is the main module that allows genuine users of credit cards to perform payment activity.

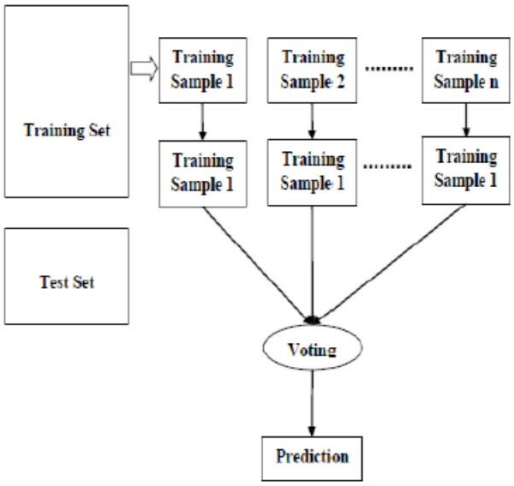

Fig 3. : RANDOM FOREST

3. MACHINE LEARNING :

The Credit card refers to card which is allocated to the customer, by allowing the customer to buy goods and services within the limit of credit card or withdraw cash in advance. The credit card gives the customer an advantage of time. Fraudulent credit card transactions are the easy targets. With no risk, amount can be withdrawn without even letting the owner known about it, within short time. The fraudsters try to make the

illegal transaction look like legitimate transaction. Hence, detecting the credit card frauds becomes very challenging and difficult.

Fraud detection methods keep on developing in order to defend the criminals in adapting to their fraudulent activities. The frauds can be classified as:

- Credit Card Frauds: Online and Offline
- Card Theft
- Account Bankruptcy
- Device Intrusion
- Application Fraud
- Telecommunication Fraud
- Counterfeit Card

PROPOSED SYSTEM :

A. Local Outlier Factor :

It is an Unsupervised Outlier Detection algorithm. 'Local Outlier Factor' refers to the anomaly score of each sample. It measures the local deviation of the sample data with respect to its neighbors. More precisely, locality is given by k-nearest neighbors, whose distance is used to estimate the local data.

B. Isolation Forest Algorithm :

The Isolation Forest 'isolates' observations by arbitrarily selecting a feature and then randomly selecting a split value between the maximum and minimum values of the designated feature. Recursive partitioning can be represented by a tree, the number of splits required to isolate a sample is equivalent to the path length root node to terminating node. The average of this path length gives a measure of normality and the decision function which we use.

Working :

This idea is difficult to implement in real life because it requires the cooperation from banks, which aren't willing to share information due to their market competition, and also due to legal reasons and protection of data of their users. Therefore, we looked up some reference papers which followed similar approaches and gathered results. As stated in one of these reference papers: "This technique was applied to a full application data set supplied by a German bank in 2006. For banking confidentiality reasons, only a summary of the results obtained is presented below. After applying this technique, the level 1 list encompasses a few cases but with a high probability of being fraudsters. All individuals mentioned in this list had their cards closed to avoid any risk due to their high-risk profile. The condition is more complex for the other list. The level 2 list is still restricted adequately to be checked on a case by case basis. Credit and collection officers considered that half of the cases in this list could be considered as suspicious fraudulent behavior. For the last list and the largest, the work is equitably heavy. Less than a third of them are suspicious. In order to maximize the time efficiency and the overhead charges, a possibility is to include a new element in the query; this element can be the five first digits of the phone numbers, the email address, and the password, for instance, those new queries can be applied to the level 2 list and level 3 list.".

RESULTS AND DISCUSSION

The Ann model detects fraud much faster than the existing system. With the help of this model, the card gets immediately blocked if found fraud is taking place. This model immediately blocks the card so the significant amount is maintained in the user's account. A security layer is added. This detection maintains a database of the users so it is helpful to submit it to the bank as proof. One can find accurate fraud detection using this technique. Credit card fraud detection can use the Ann algorithm in all payment apps (Amazon Pay, Google Pay) so that fraud can be traced. It can be used in all banking and online payment methods for secure transactions of amounts.[6]

The combination ANN technique with Machine Learning gives detailed explanation, how machine learning can be applied to fetch good results in fraud detection along with the algorithms, pseudocodes, explanation and implementation and the experiment results. With the algorithm reaches over 99.6% accuracy, the precision will remain only at 28% and This high percentage of accuracy is been expected because of the major imbalance between the number of valid and number of genuine transactions.

If both the technique of ANN and Random Forest are combined then main aim of this paper is to classify the transactions that have both the fraud and non-fraud transactions in the dataset. The process flow for the credit fraud detection problem includes the splitting of the data, model training, model deployment, and the evaluation criteria. The model have to split the data into the training data and the testing data. We use the training data to prepare the Random Forest and the Adaboost models. Then we develop both the models.

CONCLUSION

In this proposed system we have analyzed Ann's technique to detect fraud in real-time transactions. We have described Ann algorithm based on outlier detection which has less complexity and will catch fraud quickly.

Ann model algorithm has a multi-layer approach for security while the user is on way to do a transaction, as the Ann model keeps track of the user's spending way and his spending habits, patterns to detect the transaction is done by the genuine or fraudulent user.

It is concluded that a transaction was detected as a fraud when a low category does a high category payment and vice versa. As the technology is developing day by day they are also fraudsters developing. Hence it is our responsibility to get an update about the technology and use it the incorrect way. We should know the Dos and Don'ts about the credit card before we start to use it and act according to avoid any issues.

REFERENCES

[1] K. Tuyls, S. Maes, and B. Vanschoenwinkel, "Machine learning techniques for fraud detection - Google Scholar," no. May, 2000, [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Machine+learning+techniques+for+fraud+detection&btnG=.

[2] B. A. Abdulsalami, A. A. Kolawole, M. A. Ogunrinde, M. Lawal, R. A. Azeez, and A. Z. Afolabi, "Comparative Analysis of Back-propagation Neural Network and K-Means Clustering Algorithm in Fraud Detection in Online Credit Card Transactions," *Fountain J. Nat. Appl. Sci.*, vol. 8, no. 1, pp. 21–

33, 2019, doi: 10.53704/fujnas.v8i1.315.

[3] P. S. Helode, Dr. K. H. Walse, and Karande M.U., "An Online Secure Social Networking with Friend Discovery System," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 5, no. 4, pp. 8198–8205, 2017, doi: 10.15680/IJIRCCE.2017.

[4] S. C. Dubey, K. S. Mundhe, and A. A. Kadam, "Credit Card Fraud Detection using Artificial Neural Network and BackPropagation," *Proc. Int. Conf. Intell. Comput. Control Syst. ICICCS 2020*, no. Iciccs, pp. 268–273, 2020, doi: 10.1109/ICICCS48265.2020.9120957.

[5] Y. Sahin and E. Duman, "Detecting credit card fraud by ANN and logistic regression," *INISTA 2011 - 2011 Int. Symp. Innov. Intell. Syst. Appl.*, pp. 315–319, 2011, doi: 10.1109/INISTA.2011.5946108.

[6] A. U. S. Khan, N. Akhtar, and M. N. Qureshi, "Real-Time Credit-Card Fraud Detection using Artificial Neural Network Tuned by Simulated Annealing Algorithm," *Proc. Int. Conf. Recent Trends Information, Telecommun. Comput. ITC*, pp. 113–121, 2014.