ISSN(Online) : 2456-8910
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
Volume 3, Issue 6, DOI: 10.29027/IJIRASE.v3.i6.2019, 483-489, December 2019

Survey Paper for Credit Card Fraud Detection Using Data Mining Techniques

M.Deepa[1], Dr.D.Akila[2]
[1.]Ph.D. Research Scholar ,Department Of Computer Science, Vels Institute of Science, Technology And Advanced Studies (VISTAS), Chennai .Emai :deepamathan195@gmail.com
[2.] Associate Professor ,Department of Information Technology, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai
Email : akiindia.@yahoo.com

ABSTRACT

With the advent of new Technologies, nowadays electronic gadgets and online shopping's more popular. Banking and Online shopping has now become most common activities amongst the masses. As technology advances so does the risk associate with these transactions. The ease of use in this online transaction has now become more popular across the world. So it essential that we need to be very cautious on the increased Fraud activities. Online Fraud is an illegal activity that can occur when we do electronic transactions. Fraud has increased and created more risk that has serious financial loss in the financial industry. As a result, these financial institutions have enforced various techniques to improve their fraud detection methods. Since we are in the age of Information Technology, Data rules the world. So, Data mining techniques are widely used to for fraud detection. There are various algorithms such as Anomaly Detection Algorithm, Decision Tree, Random Forest, K-Nearest Neighbor, K-Means used for fraud deduction. The type of fraud doesn't remain the same in each case, so this becomes very crucial in coming up with the best algorithm for the fraudulent transaction. This paper presents the survey of those techniques and predicts the best algorithm to detect the fraudulent transaction based on a given scenario.

Keywords: *Data mining, fraud risk, finance, machine learning, Anomaly Detection Algorithm, Decision Tree, Random Forest , K-Nearest Neighbour, K-Means.*

1. INTRODUCTION

Payment industry is now providing the Digital payment method, because of its ease of use, reduce operation cost, increase market presence. People also feel it's more convenient than the tradition physical currency. There are different types of payments modes which are used in doing these transactions, such as Debit card, Credit Card, Net Banking, Wallet, UPI etc. Here we can see the ways and means to minimize the fraud in card payment systems.

Below are some of the fraud types in Card based transaction.

1) Physical Card Fraud

2) Virtual Card Fraud

Physical Card Fraud:

In most of the POS (point of sale) transaction, it's essential that the card holder has to be physically presenting the card to the merchant to carry out the transaction. There are chances that the customer card can

ISSN(Online) : 2456-8910
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
Volume 3, Issue 6, DOI: 10.29027/IJIRASE.v3.i6.2019, 483-489, December 2019

be stolen and misused by fraudsters without the customer's knowledge.

Virtual Card Fraud:

In most of the Online shopping transactions there is no need of a physical card and instead we use the Card Number, Expiry Date and CVV number to perform the transaction. This information can be stolen by the fraudsters and he can use it to perform fraudulent online transaction.

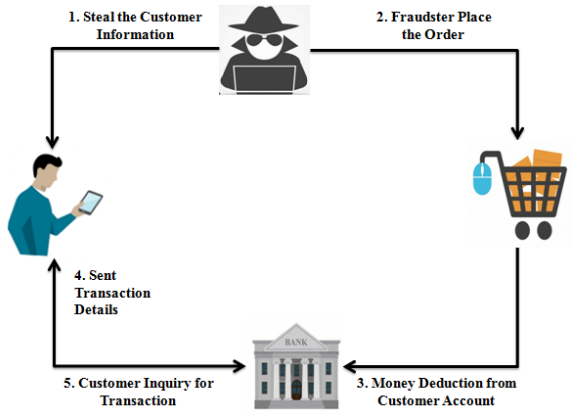

Fig:1 Transaction Flow

2. LITERATURE SURVEY IN FRAUD DETECTION

In [3] Snehal Patiletal, describes the "Decision Tree Induction Algorithm" which is used for Credit Card Fraud Detection. He describes that the decision tree approach is a new cost sensitive technique, since we eliminate the possibility by 50% while we advance to each node using the splitting attribute. The information gathered is not only used in detecting the fraudulent activity, but also get equalized and will be used in detecting the future fraudulent activity.

In [4] Dr R.Dhanapal, Gayathiri.P, uses the IP address and email for the fraud deduction using the Decision Tree technique. If the customer/merchant is suspicious, then we check by tracking if the mail id or IP address is fake.

The [2] random forest models is good on smaller datasets, however there are some problems such as imbalanced data. So there arises a need to improve the algorithm of random forest itself. In the voting mechanism, each base classifier is assigned the same weight, this may not be a right way, since the weight might not be the same for each of the base classifiers. Therefore, we also try to make some improvement for this algorithm.

Even [5] though there are so many fraud detection techniques like prediction, outlier detection, clustering etc. Machine-learning

techniques are most preferred ones in fraud detection. It's popular because of its detection rate and high accuracy. Financial institutions are interested in exploring new methods that can be cost effective.

Big bank [6] failure is now becoming most common as the effect of global financial crisis. So it's essential to determine a method which can tell us if a firm is under financial distress. The k-Nearest Neighbor classification method is the most preferred amongst the other algorithms for this. Due to critical impact it has on the stake holder "financial distress prediction model" is the buzz word in the financial research. Financial distress is the key indicator in determining the bankruptcy of a firm. The financial ratios are used in predicting financial distress. The nonparametric methods are widely used than the traditional statistical methods.

Most card fraud deduction algorithms [7] use a combination of technique like Behavior Based Technique, Genetic Algorithm and Hidden Markov Model. Here the transaction is tested individually and whatever suits the best is used for getting better result.

With the increase in the card fraudulent activity, [8] it becomes significant that we need to identify the apt machine learning algorithm and implement a combination of learning methods to come up with model which has higher accuracy in detecting the card fraudulent transactions. As we said earlier, the variables are the most crucial for the fraud determinations. Though there are many machine learning techniques, supervised machine learning is the most significant among them. Supervised machine learning can give us a right fraud score based on the scenario.

3. MOTIVATION

As technology advances so does the risk associate with these transactions. The ease of use in this online transaction has now become more popular across the world. So it essential that we need to be very cautious on the increased Fraud activities. Since we are in the age of Information Technology, Data rules the world. So, Data mining techniques are widely used to for fraud detection. This paper presents the survey of those techniques and predicts the best algorithm to

ISSN(Online) : 2456-8910
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
Volume 3, Issue 6, DOI: 10.29027/IJIRASE.v3.i6.2019, 483-489, December 2019

detect the fraudulent transaction based on a given scenario.

4. VARIOUS TECHNIQUES FOR CREDIT CARD FRAUD DETECTION

Decision Tree:

As the name indicates, Decision Tree Algorithm is used to come up with a decision based on the Tree structure. It uses Data mining induction technique that repeatedly applied on each node in the tree. The structure of a Decision tree contains a root node, multiple internal nodes and multiple leaf nodes. The nodes can represent attribute names and the edges represent the values of attributes. Any prediction here, start from the root node and traverse down till the leave based on the decision at each level by comparing the record attribute and the root attribute recursively. This is one the simplest algorithm that can be implemented and easy to understand when compared against the other algorithms[9]. Example, when tracing a mail or IP address, ore when predicting if the card holder's previous location where he has carryout the transaction, this algorithm stands as the best fit.

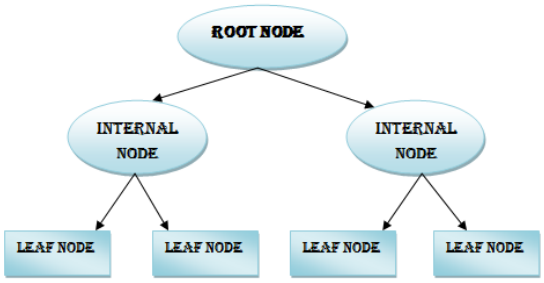

Fig:2 Decision Tree

K-means Clustering:

K-means clustering algorithm needs a group of data from the database. Clustering works on grouping the data into different cluster like high cluster, low cluster and risky cluster. It uses a pattern to get us the information needed. This can be used in finding the priority of customers and come up with similarities of fraud techniques used for the fraud detection. This algorithm used the K-means to determine if it's a legal or fraud transaction. In this transaction we declare some variables like the card number, transaction amount, transaction date, transaction origin country, customer id and merchant category code[13]. All these attributes are validated in the transaction validation section and we will store all these information in the transaction dataset. Next based on the information acquired we now classify the cluster and assign a name and label to it. Now the K-mean's algorithm will

ISSN(Online) : 2456-8910
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
Volume 3, Issue 6, DOI: 10.29027/IJIRASE.v3.i6.2019, 483-489, December 2019

decide based on the above information if it's a valid or a fraud transaction.

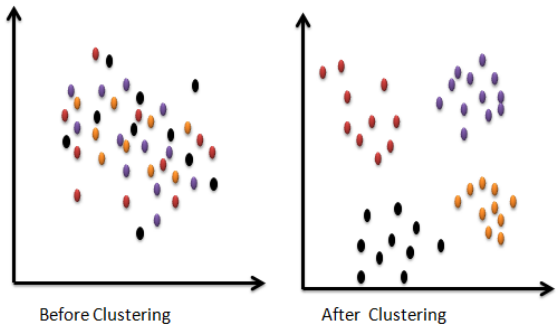

Fig:3. **K-means Clustering**

Random Forest

RFA or Random Forest Algorithm is a super set of the Decision tree algorithm. It uses multiple decision trees with supervised learning that is why it's also called as Random Decision Forest. Here we classify the observation into different categories and perform regression, so it's most effective[11]. The algorithm is more accurate when compared with other algorithms and most of our existing systems currently use this widely.

K-Nearest Neighbours

K-Nearest Neighbours algorithm uses Machine Learning along with some supervised learning. We need to give a prior data knows as the training data for this model to work[12]. This training data will classify the coordinates into different groups based on the attribute. This algorithm uses data mining, pattern recognition and intrusion detection techniques. The significant benefit in using this algorithm is this doesn't have any underlying assumptions on the data distribution[10]. The Gaussian Mixture model (GMM) in the other hand heavily uses assumptions on the given data, because of this the K-Nearest Neighbors algorithm is widely used in real-life scenarios.

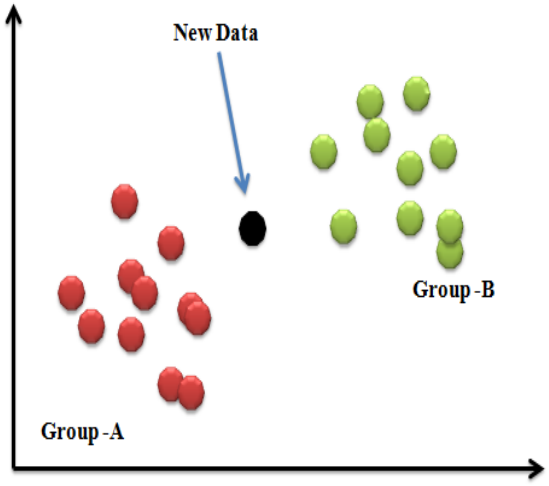

Fig:4 **K-Nearest Neighbours**

Anomaly Detection

As the name indicates, this algorithm is based on the Anomalous behaviour in a given scenario. This technique of identifying the abnormal suspicious scenarios which is

statistically different from the other observations. This anomalous behaviour is typically translated to some kind of problem like a cyber-attack or a credit card fraud or failing machine in a server etc[14].

Types of an anomaly

- ✓ **Point Anomaly:** When we map a scenario in a graph, if a point stands out far off from the rest of the data in the observation, then that point is called as Point Anomaly. This different behaviour makes it suspicious.
- ✓ **Contextual Anomaly:** When we analyse an observation, based on the context of the observation, if we find any data instance anomaly, then we term this as "Contextual Anomaly" with respect to the specific context.
- ✓ **Collective Anomaly:** When we analyse an observation, if we find a set of related data instance different than the rest of the data collection, then we term that as "Collective Anomaly".

5. CONCLUSION

Our main goal of this study is to find the illegal fraudulent transaction based on previous transaction history data. To predict the fraud score, the system uses various rules and algorithms to come up with the Fraud score for that particular transaction. Not all scenarios are the same and so why do we need to apply the same algorithm for predicting the fraud score. Instead we can apply a scenario based algorithm, which would be a best fit for that scenario.

6. REFERENCES

[1] M.Suresh Kumar, V.Soundarya ,S.Kavitha E.S.Keerthika , E.Aswini" Credit Card Fraud Detection Using Random Forest Algorithm" *3rd International Conference on Computing and Communication Technologies ICCCT 2019*, 978-1-5386-9371-1/19/$31.00_c 2019 IEEE.

[2] Shiyang Xuan, Guanjun Liu, Zhenchuan Li, Lutao Zheng, Shuo Wang, Changjun Jiang, "Random Forest for Credit Card Fraud Detection", 978–1–5386–5053–0/18/$31.00 © 2018 IEEE.

[3] SnehalPatilet al, "Credit Card Fraud Detection Using Decision Tree Induction Algorithm",International Journal of Computer Science and Mobile Computing, Vol.4 Issue.4, April- 2015.

[4] Dr R.Dhanapal1, Gayathiri.P2, "Credit Card Fraud Detection Using Decision Tree For Tracing Email And Ip", IJCSI International Journal of Computer Science, Issues, Vol. 9, Issue 5, No 2, September 2012.

[5] Rimpal R. Popat, Jayesh Chaudhary,"Survey on Credit Card Fraud Detection using Machine Learning" 2nd International Conference on Trends in Electronics and Informatics (ICOEI 2018) IEEE Conference Record: # 42666; IEEE Xplore ISBN:978-1-5386-3570-4.

[6] Sadegh Bafandeh Imandoust And Mohammad Bolandraftar, "Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background" S B Imandoust et al. Int. Journal of Engineering Research and Applications Vol. 3, Issue 5, Sep-Oct 2013, pp.605-610.

[7] Ayushi Agrawal, Shiv Kumar, Amit Kumar Mishra, "Credit Card Fraud Detection: A Case Study", *2015 2nd International Conference on Computing for Sustainable Global Development*, 978-9-3805-4416-8/15/$31.00_c 2015 IEEE

[8] *Sahil Dhankhad, Emad A. Mohammed, Behrouz Far*, "Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study", 2018 IEEE International Conference on Information Reuse and Integration for Data Science, 978-1-5386-2659-7/18/$31.00 ©2018 IEEE.

[9] **G.Suseendran**, E.Chandrasekaran "Interference Reduction Technique in Mobile Adhoc Networks Using Mathematical Prediction Filters, International Journal of Computer Applications, Volume 60, Issue.6, December 2012. pp-9-16 Doi: 10.5120/9694-0843

[10] K.Rohini**, G.Suseendran**, "Aggregated K Means Clustering and Decision Tree Algorithm for Spirometry Data", Indian Journal of Science and Technology, Volume 9, Issue 44, November 2016 pp.1-6 DOI: *10.17485/ijst/2016/v9i44/103107*

[11] R. Kiruthiga, **D.Akila, "** Phishing Websites Detection Using Machine Learning**",** International Journal of Recent Technology and Engineering, Vol.8,(2S11), September 2019, pp. 111-114 , *DOI: 10.35940/ijrte.B1017.0982S1119*

[12] C. Sudha, **D. Akila,** "Detection Of AES Algorithm for Data Security on Credit Card Transaction", International Journal of Recent Technology and Engineering (IJRTE) Volume-7, Issue-5C, February 2019,pp.283-287

[13] D.Akila, C.Jayakumar, "Acquiring Evolving Semantic Relationships for WordNet to Enhance Information Retrieval", **International Journal of Engineering and Technology**, Volume 6, November 2014

[14] Krishna PrakashKalyanathaya, **D. Akila ,**P. Rajesh , "Advances in Natural Language Processing –A Survey of Current Research Trends", International Journal of Recent Technology and Engineering (IJRTE) Development Tools and Industry Applications, Volume-7, Issue-5C, February 2019 pp.199-202.