ISSN: 0474-9030

Our Heritage

Vol-68-Issue-15

January-2020

UGC Care Listed Journal

A Study of Advance Fee Fraud Detection using Data Mining and Machine Learning Technique

Jalindar Gandal, Dr. R.G.Pawar

Abstract :
The Advance Fee Fraud is the system which initiates the global economy to develop significantly. Financial institutions are at an extraordinary risk of being used in money laundering schemes, particularly in emerging economies such as India, due to insider threat. Education and knowledge remains crucial in defending against identifying Advance Fee Fraud as the scammer techniques and tactics are continuously evolving. This survey represents an organized analysis of data mining methods and its applications in Advance Fee Fraud. This survey must be very helpful for a government officials providers to select an appropriate solution for their problem as well as for researchers to have comprehensive of the review of literature in their area

Keywords : *Victimization, Phishing, Advance Fee Scam, Artificial Neural Network, Data Mining & Machine Learning technique*

1. INTRODUCTION

In these modern days the style of payment types are changed into online transactions with special reference to banking transaction. There are several types of payment for civilizing online transaction which includes e-cash, card payments, internet banking, and e-services. Advance fee fraud one of the most conventional methods of online fraud by using internet technology. It is classified into several types .If the intended victim is interested in the deal, they are requested to forward a variety of paperwork which generally includes blank company letterheads which are duly signed, blank invoices, telephone and fax numbers, and especially bank account details. These being required to affect the transfer of the money into the bank account.

- Lottery Fraud
- Recruitment Fraud
- Friendship Fraud
- Forwarding Money Scam
- Phishing
- Pharming
- Smishing
- Identity theft
- Share Offer fraud

1.1 ADVANCE FEE FRAUD:

Advance fee fraud, more commonly referred to as Nigerian fraud or 419scam, is a prevalent form of online fraud that not only causes financial loss to individual and business but also can bring emotional or psychological damage to victim users.Advance fee fraud is a popular form of fraud in which fraudster tricks the victims into paying a certain amount of money under the promise of future. Originally the scam phenomenon started by postal mail and then evolved into business run via fax first and email later. The scam prosecution of such criminal activity is complicated, as a result, report of such crime still appear in the social media and online communities.Now a day, advance fee fraud is often perceived as a particular type of spam. Advance fee scam activities are still largely performed in manual way. More ever scammers use very primitive tools, where operations are often completely automated. Even though today advance fee fraud messages are eclipsed by large amount of spam sent by phishing, they are still a problem that causes substantial financial losses for a number of victims all around the world.

1.2 CONSEQUENCE OF ADVANCE FEE SCAM

According to 2018 global report on the financial crime scammers looted Rs.25240 Cr. ($1670 million) from Indians. India is not only home to an increasing number of victims but has also become a base for scammers targeting other countries. As much as 80 per cent of the money scammers earn in India is shared with rings outside the country, according to the global financial report. Such fraud is also becoming a source of revenue for terror groups.

1.3 APPLICATIONS OF MACHINE LEARNING

Following are the different kind of application domains where machine learning used widely

- Voice Assistants
- Amazon Recommendation System
- Google Maps
- Email Filtering
- Chatbots
- Video Surveillance
- **Fraud Detection**
- Self-Driving Cars

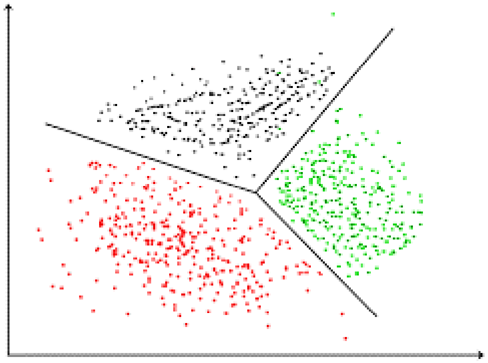
2. VARIOUS TECHNIQUES FOR ADVANCE FEE FRAUD DETECTION

In data mining there are numerous methods for identifying the advance fee fraud detection. In this Survey paper we discuss some most useful methods.

- K-Means Clustering
- Decision Tree
- Neural Network
- Hidden Markov Model
- Genetic Algorithm

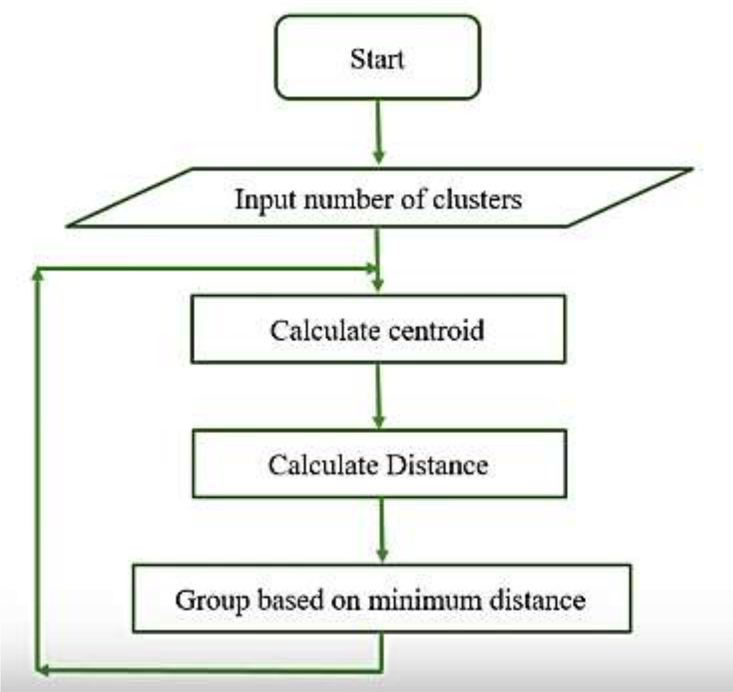
2.1 K-Means Clustering

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters).



Use of Clustering: Clustering is used in market segmentation; where we try to find customers that are similar to each other whether in terms of behaviors or attributes, image segmentation/compression; where we try to group similar regions together, document clustering based on topics,etc.

Fig: K means statistics behind



K-means algorithm is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group.

- Euclidean metric is the "ordinary" straight-line distance between two points.
- Exploratory Data Analysis Technique.
- Implement non-hierarchical method of grouping objects together.
- Determines the centroid using Euclidean Method for Distance Calculation.
- Group the objects based on minimum distance.

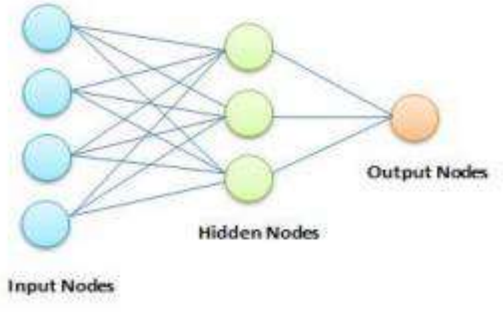
In analysis of data for recognize the same pattern or group of information clustering method can be used. It facilitates the bank to draw a decision based on the importance of client and to expose similar kind of techniques used in fraud detection. K-means clustering algorithm is most used method to identify whether the transaction is fraud or legal. transaction we declare some variable like, transaction amount, transaction date and id, country, merchant category id. Here transaction number is must. If we don't have we can't do the transaction. This process is done in transaction validation section. The information which we got as an input will going to store in transaction dataset. Next we assign the cluster name in which type of transaction that is and label it as, low cluster or high cluster or risky cluster. The transaction detail will take over to k-means algorithm. If the transaction is fraud or legal it displays a message.

2.2 Decision Tree

A Data mining induction method that recursively distributes a set of records is Decision Tree Algorithm. This is a method used for solving regression and classification problems. It used the tree representation. It contains one root node, child nodes and leaf nodes. Attribute names are used to labeled the attributes. Values of attributes are used to label the edges. For foretell a label of a class the following method is used. First, it begins from the root node then it compares the cost of the root with record node value. With this result it follows the division corresponding to that cost and travelled to the next node. This process is continued until it arrives the leaf node with predicted class value. It is simple to execute, recognize and exhibit when comparing to other classification algorithm. It is also used for tracing the mail and IP address for detecting credit card fraud. The detection depends on the location. It compares the location of preceding usage of with the present places transaction

2.3 Neural Networks

Neural network is a technique which is also used for detecting illegal usage of online functionalities given by bank with respect to scamming. It shows is valuable result in numerous problems. Neural network is a method which is based on the working principle of human brain. Like human brain, neural network also stores the existing knowledge and uses that information when needed. In detecting illegal usage of internet or online banking facilities, neural network split the information into different groups. It depends on the victims or the person who suffers advance fee scam earnings, career, and payment details frequency and counting of large purchases. This entire detail is going to evaluate the future transaction that is whether the transaction is scam or authentic. The following figure represents the process of neural network. It has three distinct types of layers.



- Input Layer: Input nodes are used to identify the victim's details and by using this information it will verify the uniqueness of the transaction.
- Hidden Layer: It perform neural network operation to discover whether the transaction is authenticated or not.
- Output Layer: After analyzing the transaction, output nodes gives the result value between 0 and 1.

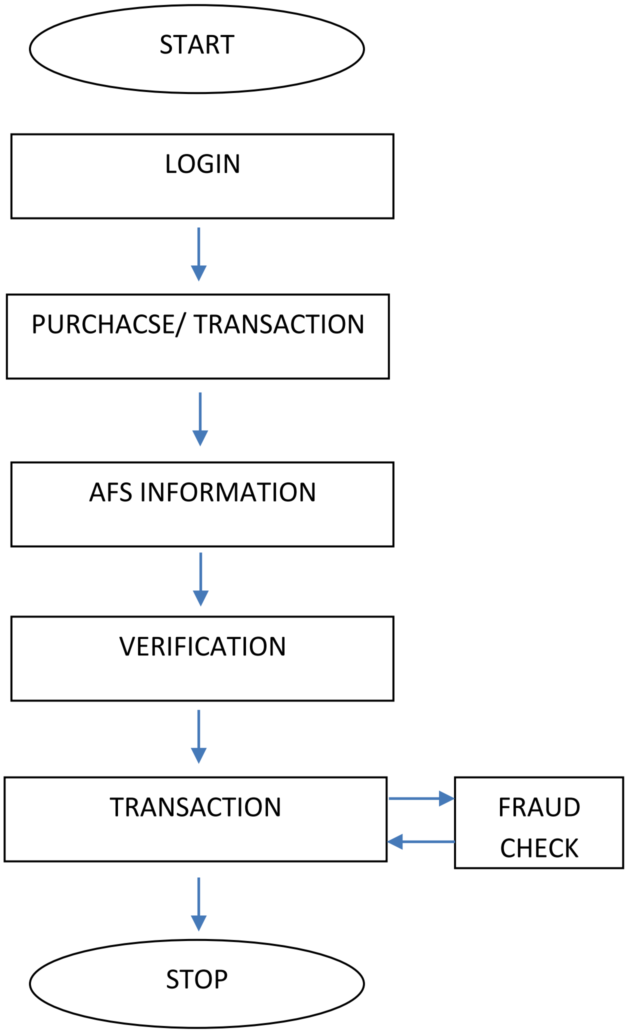
2.4 Hidden Markov Model

A set of states connected with the probability distribution is known as Hidden Markov Model. Each and every state produces an output according to the probability distribution which is based on the specific state. In this method only output can be visible to the user so that it is called as Hidden Markov Model. In detecting dishonest transaction spending pattern of the authenticated user is calculated by the previous record of transaction which has the attributes like amount that has been transferred, IP address, place of delivery and location of most recent transaction etc. The behavior of the bank user is categorized into three types. They are,

- Low spending behavior
- Medium spending behavior
- High spending behavior

The person who pay low amount for purchase are categorized into behavior of low spending. The person who spends reasonable level of amount are said to be the behavior of medium spending. And finally the person who spends huge amount is categorized into high spending behavior.

The following picture demonstrates the detection of Advance Fee Fraud using HMM.



The first level is identification of the consumer that depends on the purchasing patterns of the cardholder. It follows two step processes to identify the illegal usage of credit card. Hidden Markov Model has been prepared by using previous history of transactions. It obtains the input and validate whether the transactions details are accepted by previous training series are not.

2.5 Genetic Algorithm

To get the improved optimal solution genetic algorithm is used. It is also used to identify the fraud transactions with the given sample data set. This method is efficient and secure. It checks whether a transaction is authenticated or not. Transaction using banks has n number of attributes. At beginning it choose the data set that are going to be processed. Then we select the normalized data from the selected dataset that holds the entire detail about the person who victimized. First it calculates the critical values using regularity usage of online banking, present bank balance, overdraft and place where the particular transaction happened and average daily spending. Then it compares the data and finally it determines whether the transaction is authenticated or not.

3. CONCLUSION

This paper represents the survey on detecting illegal usage of internet banking or online banking and the strategies which concerned inside the detecting advance fee fraud. Particularly, classification and prediction mission are very essential within the technique of online banking. By way of this research the credit score card presents to discover the real clients and they can lessen the cost and additionally growth profit. They

carry a prepared evaluation, suitable time period of information-set and also appropriate choice of data set.

References

[1] Gorade, S. M., & Deo, P. A. (2017). A Study of Some Data Mining Classification Techniques, 3112–3115.

[2] Kumari, S. (2017). A Review on Various Techniques and Approaches for Credit Card Fraud Detection, 6(5), 485–489.

[3] Malini, N., & Pushpa, M. (2017). Research in Computer Applications and Robotics Analysis on Credit Card Fraud Detection Techniques By Data, 5(5), 38–45.

[4] Thomas, S. S. (2017). Implementation of Data Mining Techniques Monetary Domains. International Conference on Innovative Mechanisms for Industry Applications, (Icimia), 730–734.

[5] Gupta, A., Kumar, D., & Barve, A. (2017). Hidden Markov Model based Credit Card Fraud Detection System with Time Stamp and IP Address, 166(5), 33–37.

[6] John, S. N. N., O, O. K., Kennedy, C. G., Anele, C., Kennedy, O. O., Olajide, F., & Kennedy, C. G. (2016). Realtime Fraud Detection in the Banking Sector Using Data Mining Techniques/Algorithm. 2016 International Conference on Computational Science and Computational Intelligence (CSCI), 1186–1191. https://doi.org/10.1109/CSCI.2016.223

[7] Ashwini, S. D. K. (2017). Survey on Techniques of Data Mining and its Applications, 9359(2), 198–201.

[8] Carneiro, N., Figueira, G., & Costa, M. (2017). A data mining based system for credit-card fraud detection in e-tail. Decision Support Systems, 95, 91–101. https://doi.org/10.1016/j.dss.2017.01.002

[9] Gaur, S., Maheshwari, A., Scholar, M. T., Dhruwa, L., & Upadhyay, A. (2017). Hidden Markov Model and Genetic Algorithm Based Credit Card Fraud Detection, 4(6), 565–577.

[10] Kajaree, D., & Behera, R. . (2017). A Survey on Machine Learning: Concept, Algorithms and Applications. International Journal of Innovative Research in Computer and Communication Engineering, 5(2), 1302–1309. https://doi.org/10.15680/IJIRCCE.2017.