International Journal of Advanced Scientific Research | IJASR
International Journal of Scientific Research in Technology, Applied Sciences & Health Studies | IJSRTASHS
ISSN Print: 2579-1036 | ISSN Online: 2579-1044
Volume 3, Number 1, June, 2018

Data Mining Application in Credit Card Fraud Detection System

[1]Amanze, B. C., & [2]Inyiama, H. C.
[1]*Department of Computer Science,*
Faculty of Science, Imo State University, Owerri.
[2] *Department of Computer Science,*
Nnamdi Azikiwe University, Awka.

Abstract

Since the evolution of the internet, many small and large companies have moved their businesses to the internet to provide services to customers worldwide. Credit-card fraud is increasingly rampant in the recent years for the reason that the credit-card is majorly used to request payments by these companies on the internet. Therefore the need to ensure secured transactions for credit-card owners when consuming their credit cards to make electronic payments for goods and services provided on the internet is a criterion. Data mining has popularly gained recognition in combating credit-card fraud because of its effective, and machine learning algorithms that can be implemented to detect or predict fraud through Knowledge Discovery from unusual patterns derived from gathered data. This system implements the unsupervised anomaly detection algorithm of data mining to detect fraud in a real time transaction on the internet, and thereby classifying the transaction as legitimate, suspicious fraud and illegitimate transaction. The anomaly detection algorithm is designed on the data mining technique which implements the working principal of the human brain. To understand how credit card fraud are being committed, in this study the different types of fraudsters that commit online credit card fraud and the techniques used by these online fraudsters to commit fraud on the internet is discussed.

Keywords*: Credit card fraud, Fraudsters, Data mining*

Corresponding Author: Amanze, B. C.

Background to the Study

Currently, data mining is a popular way to combat frauds because of its effectiveness. Data mining is a well-defined procedure that takes data as input and produces output in the form of patterns (Hand *et al.,* 2002). In other words, the task of data mining is to analyze a massive amount of data and to extract some usable information that one can interact with for future uses. Once one has the right model for the data, the model can be used for predicting future events by classifying the data. In terms of data mining, fraud detection can be understood as the classification of the data. Input data is analyzed with the appropriate model that determines whether it implies any fraudulent activities or not.

A well-defined classification model is developed by recognizing the patterns of former fraudulent behaviors. Then the model can be used to predict any suspicious activities implied in a new data set. Data mining and model construction require a lot of time, which prohibits it to detect frauds in real time. This is a serious drawback since, in many occasions such as online credit card transactions, one need to detect fraudulent activities in a very short period of time, typically while the fraudster is still at the banking hall. Otherwise, the loss could be huge.

Data mining refers to a family of machine learning techniques capable to analyze and extract non-trivial patterns from data (Chen *et al*., 2006). Data mining is also known as knowledge discovery because it can reveal previously unknown information which was hidden in the data of various databases. The mined information can be proved very useful for the organizations who apply data mining. Based on the results, organizations may make important decisions which can help them survive in the competitive environment. For instance an organization can analyze the sale records of its customers in order to send attractive offers on the most popular products. Hormozi and Giles (2004) state that "data mining enables an organization to focus on the most important information in the database, which allows managers to make more knowledge decisions by predicting future trends and behaviors (Hormozi and Giles, 2004). Given that databases are too large; it is very inconvenient and impractical to look manually for hidden patterns on the data. Therefore data mining can be introduced to facilitate the discovery of useful knowledge. Forrester Research firm reported that 52%, of 1000 companies in total, decided to employ data mining techniques in 2001 to improve their marketing strategies; an increase of 34% comparing to 1999 (Hormozi and Giles, 2004).

Data mining can also be used to detect fraudulent credit card transactions, predict which customers are more likely to default their contractual obligations by going bankrupt as well as identify fraudulent credit applications. Srivastava *et al.* (2008) state that the only way to detect credit card fraud is by analyzing the spending behavior of customers using data mining techniques (Kundu *et al.* 2004). Customers tend to follow a standard spending profile and therefore any transaction which deviates from that standard can be considered as suspicious. Suspicious transactions can be examined in detail by bank officers to determine whether they are indeed fraudulent or not. Like most of the machine learning algorithms, data mining techniques tend to learn models from data. Data mining

is becoming a strategically important area for many business organizations including the banking sector. It is a process of analyzing the data from various perspectives and summarizing it into valuable information. Data mining assists the banks to detect pattern in a group and unearth unknown relationship in the data. It can also be applied in credit risk management and fraud detection. In credit risk management, banks provide credit card to its customers by verifying the various details. Even though, banks are cautious while providing credit card, there are chances for credit card defaults by customers. Data mining technique can help to distinguish customers who collect credit card promptly from those who do not. It also helps to predict when the customer is at default, whether providing credit card to a particular customer will result in bad credit card etc. The data mining tool will compile all data bank-wide and construct several new rules to detect fraud. It also flags the branch where these problems occur.

Data mining technique also helps to analyze such pattern as behavior and reliability of the customers, the given demographics and transaction history of the customers and transactions that can lead to fraud. Since Banks record all installment credit cards on a central mainframe system, these files contain all the essential information pertaining to a credit card. This data serve as basis on which the search for any irregularities in the credit card process begins. With data mining, a thorough profiling and ranking of banks with respect to credit card fraud risk can be achieved. The real power of the data mining tool lies in the fact that it has ability to connect data we believed was uncorrelated, and construct new rules. Fraud rule results are converted into risk scores and then displayed by the reporting application. The reporting application gathers all the information from the rules and transforms these absolute numbers in percentages and relative scores. This data is data is then combined to create total risk scores for each bank countrywide. The higher this score, the more likely irregularities occurred at that specific bank. The reporting tool allows the user to choose from an extensive array of graphs to plot these risk scores. “That way, one can instantly spot where irregularities occur, determine whether these are intentional or not, and take appropriate action in case of fraud. With the reporting tool, inspectors can pin-point very specific internal fraud irregularities. Inspections can now be carried out in much less time with increased accuracy and efficiency. There are three approaches on learning the data mining models.

Credit Card Hacking

In 1980 the term hacking became a buzzword, which was taken to be derogatory and by the misuse or overuse was attached to any form of socially non-acceptable computing activity outside of polite society. Credit card hacking is harder to do using traditional methods such as decrypting the magnetic stripes and recreating them. Hackers were assumed to be the fringe society of the computing fraternity, who did not know any better and who had obtained access to a technology with which they terrorized the world of communications and computing. These connotations are in contrast to the use of the term in the 1950's and 1960's when hackers were at least to be tolerated for their potential, though not necessarily displayed in public. Scientists such as Edison (electric light bulb, phonograph, etc.), Fleming (penicillin), Barnes-Wallis (the bouncing bomb and swept

wing aircraft), Watson-Watt (radar) and possibly even Babbage (the difference and analytical engines), may have been honored to be identified as hackers. Only in more recent times has there been confusion between the terms hacker, petty criminal (Lee *et al.*, 2011).

The concept of hacking as a methodology to achieve some particular goal has the allusion of working at something by experimentation or empirical means, learning about the process under review or development by ad hoc mechanisms. In hacking a computer, the enhancement of the system is an end in itself. Applications of that system don't count. In the same manner, there is not any particular way or any life cycle to do hacking and there is no specific end goal, an improvement is in itself an achievement, but not necessarily a reason for further activity. While hacking was generally counter-society it is not necessarily anti-society (Lee *et al.*, 2011).

Different Ways of Hacking

In an online credit card purchase, the payment data transfers between the customers PC and the vendor's shop over the internet. That raises concerns about credit card online security and identity theft. Most online shops are secured to prevent unauthorized people from seeing that information and you should see a secure site symbol displayed by the Web browser as proof. If one doesn't see evidence of a secure site, the transfer of personal information, including the credit card data, could be exposed and subject to theft. One should be careful and should think before entering the data. Because it is difficult to intercept information transferred over a secure connection (FSPro-Labs, 2011).

Different Types of Fraudster

Fraudsters usually fall into one of three categories:

1. Pre-planned fraudsters, who start out from the beginning intending to commit fraud. These can be short-term players, like many who use stolen credit cards or false social security numbers; or can be longer-term, like bankruptcy fraudsters and those who execute complex money laundering schemes.
2. Intermediate fraudsters, who start off honest but turn to fraud when times get hard or when life events, such as irritation at being passed over for promotion or the need to pay for care for a family member, change the normal mode,.
3. Slippery-slope fraudsters, who simply carry on trading even when, objectively, they are not in a position to pay their debts, this can apply to ordinary traders or to major business people (Wells, 2016).

In 2007, Klynveld Peat Main Goerdeler (KPMG) carried out research on the Profile of a Fraudster (KPMG survey), using details of fraud cases in Europe, India, the Middle East and South Africa. The Association Certified Fraud Examiners (ACFE) carried out similar research on frauds committed in the US.

These surveys highlight the following facts in relations to fraudsters:

a. Perpetrators are typically college educated white male
b. Most fraudsters are aged between 36 and 55
c. The majority of frauds are committed by men
d. Median losses caused by men are twice as great as those caused by women
e. A high percentage of frauds are committed by senior management (including owners and executives)
f. Losses caused by managers are generally more than double those caused by employees
g. Average losses caused by owners and executives are nearly 12 times those of employees
h. Longer term employees tend to commit much larger frauds
i. Fraudsters most often work in the finance department, operations/sales or as the CEO.

The ACFE report also found that the type of person committing the offence depends on the nature of the fraud being perpetrated. Employees are most likely to be involved in asset misappropriation, whereas owners and executives are responsible for the majority of financial statement frauds. Of the employees, the highest percentage of schemes involved those in the accounting department. These employees are responsible for processing and recording the organization's financial transactions and so often have the greatest access to its financial assets and more opportunity to conceal the fraud.

Who are the credit-card fraudsters?

I. **Credit-card information buyers***:* They are fraudsters with little or no professional computer skills (e.g. Computer Programming, Networking, etc.) who buy hacked (or stolen) credit-card information on an illegal "credit-card sales" website. They buy this credit-card information with the intention of making electronic payment for some good and services on the internet (John, 2013).

II. **Black hat hackers**: Recent research on Hackers in terms of Computer Security defined a "black hat hacker" (also known as a cracker) as a hacker who violates computer security with malicious intent or for personal gain **8**. They choose their targets using a two-pronged process known as the "pre-hacking stage"; Targeting, Research and Information Gathering, and Finishing the Attack. These types of hackers are highly skilled in Computer Programming and Computer Networking and with such skills can intrude a network of computers. The main purpose of their act of intrusion or hacking is to steal personal or private information (such as credit-card information, bank-account information, etc.) for their own personal gain (for instance creating a "credit-card sales" website where other cyber credit-card fraudsters with little or no computer skills can buy credit-card information).

III. **Physical credit-card stealers**: They are the type of fraudsters who physically steal credit-cards and write out the information on them. They physically steal these

plastic credit-cards (maybe by pick-pocketing in a crowded place) and write out the credit-card's information with the intention of using this credit-card information to make electronic payment for some good and services on the internet.

Fraud Detection in Banking Sector

Sometimes the given demographics and transaction history of the customers are likely to defraud the bank. Data mining technique helps to analyze such patterns and transactions that lead to fraud. Banking sector gives more effort for Fraud Detection. Fraud management is a knowledge-intensive activity. It is so important in fraud detection is that finding which ones of the transactions are not ones the user would be doing.

1. **The Clustering model:** Clustering helps in grouping the data into similar clusters that helps in uncomplicated retrieval of data. Cluster analysis is a technique for breaking data down into related components in such a way that patterns and order becomes visible. This model is based on the use of the parameters' data cauterization regions.

In order to determine these regions of cauterization first its need to find the maximum difference (DIFF max) between values of an attribute in the training data. This difference ($DIFF_{max}$) is split into $N_{interval}$ segments. $N_{interval}$ is the binary logarithm of the attribute values account N_{point}. In general, $N_{interval}$ can be found using another way of looking such calculation of $N_{interval}$ is based on the assumption that a twofold increase of N_{points} will be equal to $N_{interval}$ plus one.

Thus $N_{interval}$ centres and corresponding deviations that describe all values of the certain attribute from the training data appears. The final result of classification of the whole transaction is the linear combination of classification results for each parameter:

$$\text{Result} = w1 \times \text{Class1} + \ldots + wn \times \text{Class n} \qquad (1)$$

2. **Probability density estimation method:** To model the probability density function, Gaussian mixture model is used, which is a sum of weighted component densities of Gaussian form.

$$p(x) = \sum_{j=1}^{M} p(x \mid j)\, p(j) \qquad (2)$$

The $p(x \mid j)$ is the jth component density of Gaussian form and the P (j) is its mixing proportion. The parameters of the Gaussian mixture model can be estimated using the Expectation Maximization (EM) algorithm computes maximum-likelihood estimates of parameters .The on-line version of the EM algorithm was first introduced by Nowlan.

$$P(j)_{new} = \alpha P(j)_{old} + P(j \mid x). \qquad (3)$$

Remembering that the new maximum likelihood estimate for P (j) is computed as the expected value of P (j | x) over the whole data set with the current parameter fit.

Methodology

Data mining methodology will help the bank to identify who uses credit card to commit fraud. It also helps to predict when the card is at default, and it uses the credit card to determining and analyzes the behavior and reliability of the customers. With data mining techniques, banks can do a thorough profiling and ranking of their branches with respect to credit card fraud risk. To accomplish this, relevant information can be gathered from the credit risk information service databases. These files contain all the essential information pertaining to a credit card fraud. That includes characteristics such as identity of cardholders, location of the branch/bank where the credit card was issued and where the changes are made to the credit card.

Conclusion

Adaptive data mining and intelligent agents can play an important role in the credit card fraud detection domain. A conceptual framework for a multi-agent system based on Credit Card Fraud (CCF) process is developed, in which various classes of intelligent agents are proposed to provide a set of functionalities for CCF in electronic transaction environment for banks. They are robust enough to defeat sophisticated fraudsters, they are fast enough to minimize fraud damages, and they are scalable enough to tackle huge volumes of data. Intelligent agents will eventually be the ultimate means to fight against credit card frauds. However, there is still a long way to go before the wide adoption of intelligent agents for credit card fraud detection. The accuracy of fraud detection needs to be improved; the reliability of the agents needs to be ensured by testing the system on a real bank server to check its performance and acceptability. The study resulted in a model, which is used to detect abrupt changes in established patterns and recognize typical usage patterns of fraud. The credit card fraud detection system was designed to run at the background of existing banking software and attempt to discover illegitimate transaction entering on real-time basis. This proved to be very efficient method of discovering fraudulent transactions. The data used was a mixture of normal and fraudulent with an unknown mixing mechanism. Credit card fraud detection system detected most of the fraudulent transactions. The research presented the following contributions to knowledge –the design of intelligent agent credit card fraud (CCF) detection model using multi-agent; design of the CCF detection agent architecture and design of the CCF detection alert.

An approach is used to develop the credit card fraud detection system that utilizes both adaptive data mining and intelligent agents' approaches to achieve a synergy that better handles the Nigerian credit card fraud situation using multi-agents instead of two-stage model normally used in fraud detection algorithm. This reduced the classification of legitimate transactions as fraudulent, ensured accurate and reliable result. The study reinforces the validity and efficiency of data mining and intelligent agent as a research tool and laid a solid groundwork for intelligent detection methodologies to be used in an operational fraud detection system.

Recommendation

It is recommended that all banking sectors should integrate the model developed in this dissertation into their banking software so as to help detect credit card frauds. More also credit card users are advised to maintain secrecy about their credit card pins and other details. This will help eliminated credit card fraud and encourage cashless transactions.

References

Chen, R. C., Chen, T. S., & Lin, C. C. (2006). Detecting credit card fraud by using questionnaire, responded transaction model based on support vector machines. *Springer-Verlag Berlin Heidelberg,* 800-806.

Hormozi, A. M & Giles, S. (2004). Data mining: A competitive weapon for banking and retail industries. *Information Systems Management,* 62-71.

Abhinav, S., Kundu, S., Shamik, S., & Arun, K. (2008). Credit card fraud detection using Hidden Markov Model. *IEEE transactions on dependable and secure computing*, *5*(1), 83-86

Lee, W., Stolfo, S., & Mok, K. (2011). Adaptive intrusion detection: a Data mining approach*, Kulwer Academic Publishers*.

Wells, F. (2016). *Customers over fraudulent accounts.*

John, A. (2013). Data mining application for cyber credit-card fraud detection system. *Proceedings of the World Congress on Engineering,* (3)3-5*, London, U.K.*

Liu, K., Sun, L., Dix, A., & Narasipuram, M. (2001). Norm based agency for designing collaborate information systems. *Information Systems Journal*, 11(3), 229-47.

Madey, G., Freeh, V., & Tynan, R. (2002). The open source software development phenomenon an analysis based on social network theory. *Proceedings of the 8th Americas Conference on Information Systems*, *AMCIS, Dallas, 1806-1813.*

Srivastava, J., &Raghubir, P. (2008). Monopoly Money, the Effect of Payment Coupling and form on spending behavior. *Journal of Experimental Psychology, 27(4), 460-474* SPro-Labs 2011