A

Project Report

On

Online Fraud Detection System on logistic Regression

Submitted in partial fulfillment of the requirements

Of University of Mumbai for the degree of

Bachelor of Engineering

Ashitosh Prakash Gole. (118CP3201A)

Nitesh Prabhu Kharatmol. (118CP33031A)

For Subject

Major Project

Under Supervision

of

Prof. Sonali Patil

MAHATMA GANDHI MISSION

Department of Computer Engineering Mahatma Gandhi

Mission's College of Engineering &

Technology Kamothe, Navi Mumbai – 400 209

University of Mumbai

Academic Year: 2020-21

CERTIFICATE

This is to certify that the Project work "**Online Fraud detection using Logistic regression**" done by " **Ashitosh Prakash Gole**, **Nitesh Prabhu Kharatmol** " students of "Department of Computer Engineering" is a record of bonafide work carried out of them. This Project is done as partial fulfilment of obtaining "Bachelor of Computer Engineering BE Sem VII" degree to be awarded by "Mahatma Gandhi Mission of College of Engineering and Technology, Kamothe". The matter embodied in this project report has not been submitted to any other university for the award of any other degree.

Prof. Sonali Patil	Prof. Vijay Bhosale
(Supervisor/Guide)	(Head of Department)

Date:- _____

DECLARATION

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

- -

(Ashitosh Prakash Gole)

(Nitesh Prabhu Kharatmol)

Date: - _____

ACKNOWLEDGEMENT

We would like to express deepest appreciation towards Dr. Geeta Lathkar, Director, Mahatma Gandhi Mission's College of Engineering and Technology, Dr.Vijay R. Bhosale, Head of Department of Computer Engineering whose invaluable guidance supported us in completing this project.

We are profoundly grateful to Prof. Sonali Patil for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion.

At last, we must express our sincere heartfelt gratitude to all the staff members of Computer Engineering Department who helped us directly or indirectly during this course of work.

TABLE CONTENTS

1.0 ABSTRACT

Transaction fraud imposes serious threats to e-commerce shopping. As the online transaction is becoming more well known the types of online transaction frauds associated with this are likewise rising which affects the money related industry. This fraud detection system has the ability to restrict and hinder the transaction performed by the attacker from a genuine user's credit card details. To overcome these problems, this system here is developed for the transactions higher than the customer's current transaction limit. During registration, we take the required data which is efficient to detect fraudulent user action. The details of items purchased by any Individual transaction are generally not known to any Fraud Detection System (FDS) running at the bank that issues credit cards to the cardholders. BLA (Behavior and Location Analysis) is implemented for addressing this problem. A FDS runs at a credit card giving bank. Each approaching transaction is submitted to the FDS for verification. FDS receives the card details and transaction value to verify, whether the transaction is genuine or not. The types of products that are purchased in that transaction are not known to the FDS. The bank declines the transaction if FDS affirms the transaction to be a fraud. User spending patterns and geographical area is used to verify the identity. In the event that any surprising pattern is detected, the system requires re-verification. Based on previous information of that user, the system recognizes uncommon patterns in the payment procedure. After 3 invalid attempts, the system will hinder the user.

Keywords – Credit card fraud detection, fraud detection techniques, E-commerce

2.0 Introduction

Information mining involves the use of complicated information investigation instruments to discover previously obscure, substantial patterns and relationships among large informational indexes. These apparatuses can include mathematical calculations, factual models, and machine learning methods, (for example, Neural Networks or Decision Trees). Consequently, information mining comprises more than the collection and management of information, it likewise includes investigation and prediction. Information mining can be performed on information represented in textual, quantitative or multimedia structures. Information mining applications can use a range of parameters to observe the information. This includes an affiliation, characterization, sequence or way examination, clustering, and forecasting. When utilizing typical measures, detection of credit card fraud is a dubious errand. Therefore, the development of the credit card fraud detection model has become a lot of significant, whether in the academic, association, or business network recently. Proposed models or existing models are generally insights driven or Artificial Intelligent-based (AI), which have the theoretical advantages in not forcing counterfeit suspicions on the information variables [1].

Credit Cardholders have numerous beneficiary schemes to hold interest-free balances for just about two months with "grace-period". Suitable data on fraudulent activities is strategic to the financial business. Banks have huge databases. The extraction of significant business data should be possible from these information stores. Datastores have some patterns into clusters that are normal to the information. The concept of fraud detection has been laid on information mining techniques which include affiliation rules, clustering, and order. The chief purpose of research on fraud detection has been focused on pattern coordinating in which irregular patterns are identified from the typical ones. The prevalence of online shopping is developing step by step on high pace. These days, Credit card is the most mainstream mode of payment (59 percent), Germany and Great Britain have the largest number of online shoppers. For the most part, Retailers like Wal-Mart handle a lot larger number of credit card transactions online just as regular purchases [2]. There are numerous choices for taking care of credit card payments on the Internet, as the processing of credit card transactions is generally independent of the type of e-commerce exchange. While a huge segment of e-commerce would comprise of credit card purchases, like regular or often. It is more significant for businesses and associations that rely upon on income from e-commerce to realize the alternatives available just as costs linked with credit card transaction processing on the Internet. Nobody has any clue about the transaction being processed are whether a fraudulent transaction or legitimate which has passed the prevention mechanisms. Therefore,

the objective of the fraud detection system is to pre-determine every transaction for the chance of being fraudulent regardless of the prevention mechanisms and to categorize transactions as fraudulent ones as early as possible after the fraudster has begun to submit a fraudulent transaction. Credit card fraud detection is a tremendous errand yet in addition trendy problem to solve. Numerous fraud detection systems estimate the transactions and generate a doubt score (generally a likelihood between 0 and 1) which demonstrates the chances of that transaction to be fraudulent. Computational procedures of these scores are applicable to the techniques used to construct the model(s) in the fraud detection systems. These corresponding scores are used with a predefined threshold value to differentiate between fraudulent transactions from the legitimate ones easily [3].
Presentation of new technologies, for example, telephone, automated teller machines (ATMs) and credit card systems have enlarged the measure of fraud misfortune for some banks. Breaking down whether each transaction being processed is legitimate or not is very expensive is another undertaking to determine transaction genuinely. Further, on the off chance that we check them in all transactions and affirm whether a transaction was done by a client or a fraudster by calling all cardholders is cost-prohibitive. Fraud prevention via programmed fraud detections mechanism can be applied where the well-known arrangement methods can be identified, where pattern recognition systems have key capacities. One can learn from fraud that happened previously and categorize new transactions easily. Recently, perhaps the most frequently used technique is Neural Networks in the credit card business.

Credit Card Fraud Detection space presents a number of challenging issues for information mining also:

There are a large number of credit card transactions processed each day. The mining of such massive measures of information requires profoundly efficient techniques that scale information efficiently.

Highly skewed-information,

Each transaction record has a different dollar sum and there is a chance of variable potential misfortune.

2.1 The Fraud Detection Problem

The problem of detecting fraudulent transactions happens after they have been focused on fraud prevention methods and relevant processes. There is an immense literature on a wide range of security methods to take care of transactions from

unauthorized use or exposure of their private/secure data and consequent valuable resources. In any case, fraudsters discover a mode through which numerous clever means of circumventing countless prevention techniques.

On the other side, numerous transaction media, for example, ATM, bank cards, or debit cards, require the use of pins, passwords, and in some cases "biometrics" to authenticate the legitimate owner. Credit cards create intriguing problems since generally no pin is required for their use; just the name, expiration date, and the record number are required. Famous means of criminally executing with credit cards is by stealing someone's identity and in some cases, creating a new fake identity. Therefore, fraudulent electronic transactions (E-transaction) with credit card are the key problem. Credit cards need not be necessarily truly obtainable to execute and over the internet, they can be used to fraudulently execute web better and heavier losses for banks and their customers whenever got by fraudsters. The chief idea in fraud detection is that fraud might be detected by seeing huge deviation from the "ordinary behavior" of a customer's record. That is the reason; behavior of a record would thus be able to be used to protect that account. Currently, banks have come to realize that a fused, worldwide methodology is required to detect fraud, including the periodic offering to each other of data about assaults.

2.3. Different Type of Fraud Techniques

There are numerous manners by which fraudsters bring out credit card fraud. As the technology changes, so do the technology of fraudsters varies and hence the mode wherein fraudsters approach completing fraudulent activities. Frauds can be comprehensively categorized into three stages i.e., customary card related frauds, merchant related frauds, and Internet frauds. Different types of methods for submitting credit card frauds are:

A. Merchant Connected Frauds (MCF)

Merchant connected frauds are being committed either by owners of the merchant firm or their employees. Different types of frauds initiated by merchants are:

Merchant Collusion: When merchant owners or their employees intend to submit fraud utilizing the cardholder accounts or by utilizing personal data [4].

Triangulation: Triangulation is among the type of fraud which is done and operates from a website. Triangulation includes items or merchandise that are offered at heavily discounted rates and are being shipped before payment. The phenomenon initiates by the customer while perusing the site and on the off chance that he/she

likes the item he/she place the online data, for example, name, address, and legitimate credit card details to that specific site. However, when the fraudsters get these details, they order products from a legitimate site utilizing stolen credit card details. Further, after this, fraudsters use credit card data for buying the items/merchandise.

3.0 Literature Review Overview on Credit Card

Credit card frequently used as a necessary mode of payments in the present society. People used credit card for a range of reason, for example, getting credit office, loan, easy payment, charge card. There are some controversial issues that have been addressed not just in terms of the numbers of credit flooding the country's economy, however the sum transactions that end up with payment default and the numbers of credit card fraud as been recorded which endangered the economy ought to be seriously focusing [5]. But since of the advances and changing behavior in buying activities has considerably contributed to the dissemination of credit card as becoming more noteworthy and applicable in keeping up the buying activities. Based on the judgment, it is stated that there is positive connection between usage rate and income. The way that was frequently stated, a large portion of the card issuers ordinarily allowance a higher credit limit among the higher income gathering. In conclusion, it was stated that higher income clients are the primary targets for the credit card issuers. Massive purchase permits people not to convey money and is useful in Internet purchases and rental collateral. In any case, the emergency is that it is improper on religious grounds because there will be an interest payment made when the extraordinary balance isn't repaying in full.

In the card issuer's perspective, numerous problems occurred. Industry is developing and this research would be helpful for the banks offering the credit cards to concentrate on quite a few factors that pressurize the credit card holders in picking their preferred credit cards.

Butta Fogo began the workshop with an operational definition of credit card fraud as: "Unauthorized record movement by a person for which that record was not planned. Operationally, this is an event for which move can be made to stop the neglect in progress and incorporate hazard management practices to protect against comparative activities in the future." He then described the range of fraudulent activities observed in the business.

The Internet and the equivocalness associated with card not present (CNP) transactions current unique fraud management challenges. Authentication of the cardholder is an essential requirement in overseeing fraud on the Internet. There are no normally accepted arrangements. As a result, credit card fraud on the Internet is

altogether greater than in the physical, or even, phone environments [6].

Information mining contributed towards fraud detection. Information mining has different categories through which different operations has been performed. Information Mining can be for the most part classified into the accompanying categories:

Association rule mining which uncovers interesting affiliation patterns among a large set of information items by indicating attribute-value circumstances that happen together regularly. Market basket investigation is a great example wherein dissecting buying propensities for customers by discovering relationship between different items in customers' "shopping baskets. “Classification and prediction is the process of identifying a set of common features and models that to explain and recognize classes or concepts. Models are used to guess the class of objects whose class label is obscure. For example, Bank which may group an advance application as either a fraud or a potential business utilizing models based on uniqueness of the candidate. A huge number of order models have been developed for predicting future trends of financial exchange indices and foreign exchange rates (FRI).
Clustering investigation segments a massive set of information into subsets or clusters. In this, each cluster is a collection of information objects that are like one another inside the same cluster yet not at all like objects in other clusters. Moreover, objects are clustered based on the principle of augmenting the intra-class resemblance while limiting the inter-class comparability. For example, clustering techniques can be used to recognize stable dependencies for chance management just as investment management.
Sequential pattern and time-series digging searches for patterns where one value leads to another later value. Example, after the swelling rate increases, the financial exchange is likely to go down.

4. PROPOSAL WORK

Credit card fraud detection has become an inevitable piece of E-commerce applications. As the applications involve monetary transactions, fraud detection techniques are indispensable. Parcel of research has been made in this area. Numerous calculations came into existence to detect credit card fraud. The calculations include fluffy rationale, sequence alignment calculation, information mining techniques, machine learning and man-made brainpower approaches. There are other techniques, for example, Web Services – Based CCFD, CCFD with Artificial Immune System, Card watch, Bayesian Belief Networks, Intrusion Detection, Case Based Reasoning for CCFD , Advanced Fraud Detection , CCFD based on computational intelligence , CCFD utilizing self-sorting out maps. Many are based on pattern coordinating, Meta learning and man-made consciousness. In this paper we compare some of the techniques which are useful for CCFD.

5. System Design

5.1 Examination OF CCFD SYSTEMS

In this section we compare different credit card fraud detection techniques, for example, Dempster–Shafer theory and Bayesian learning, Hidden Markov Model (HMM) and Genetic Programming (GP). Dempster–Shafer theory and Bayesian learning approach The Dempster–Shafer theory and Bayesian learning approach is the blend of two approaches [17], [18], The evidences from past and correct are combined so as to detect fraud. Data combination is the methodology followed by this cross breed technique. Figure 1 shows the overview of this mixture approach. As seen in figure 1, this methodology has four components. They are rule-based filter, Bayesian learner, transaction history database, and Dempster-Shafer adder. The evidence found from multiple components are fused and the detection is made. This methodology is more accurate yet consumes more resources besides being moderate.

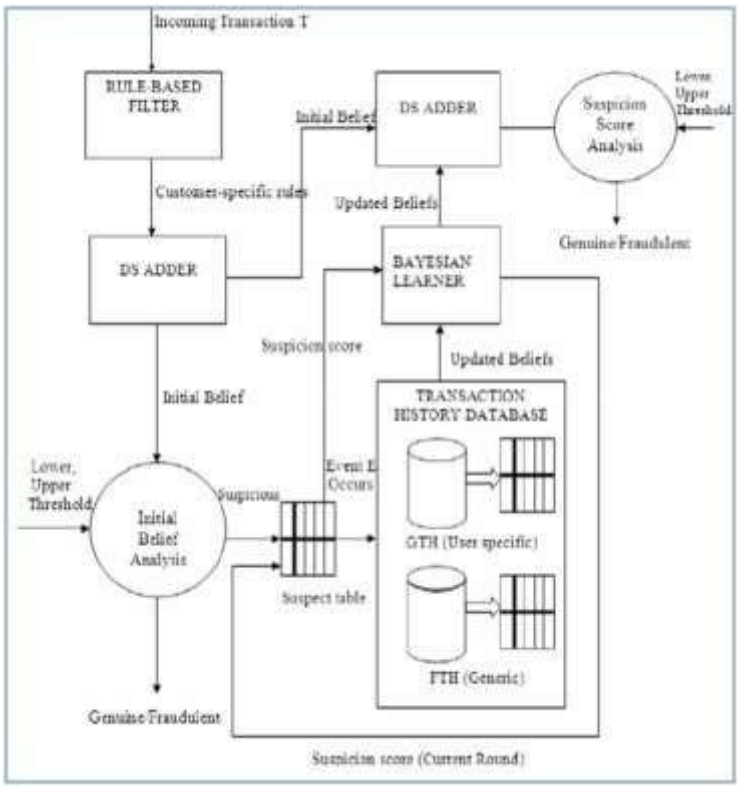

Fig. 1 – Hybrid approach for CCFD (excerpt from [16])

5.2 CCFD utilizing Hidden Markov Model (HMM)

This model is used to analyze credit card transactions. It needs preparing information and furthermore tests information so as to detect fraud. It uses the K-means information mining calculation internally. The K-means calculation takes all credit card transactions

what's more, a number of clusters as info and generates clusters that are used in HMM. All the transactions are divided into low, medium, and high, the three clusters. Once the clusters are formed they are kept in a HMM. The HMM is used for every new transaction. The sum in new transactions ought to belong to either low or medium or high. If not the transaction is suspected to be fraudulent and the corresponding people or associations are alerted. The general overview of HMM is as demonstrated as follows.

CCFD using GP

Genetic writing computer programs is widely used for taking care of different problems. In this paper, we implemented a Genetic Algorithm to detect credit card fraud. This calculation makes use of existing transactions of credit cards. It uses multiple criteria to detect fraud. The criteria include credit card usage frequency, credit card usage area, overdraft on the credit card, and credit card book balance. The overview of the GP approach is as appeared in figure 3.

As can be seen in figure 3, the data of credit cards is taken from data warehouse. Then the data is subjected to rules engine. The rules engine contains fraud case rules. The filter and priority components take care of filtering and priority setting. The genetic algorithm is responsible to detect fraud.

6.PROTOTYPE IMPLEMENTATION

A prototype application has been based on the Java stage. The application is developed with Graphical User Interface (GUI) to be user-friendly. The environment used to fabricate the application includes a PC with 4GB RAM, Core 2 double processor running Windows XP operating system. NetBeans is used as an IDE. The significant application screens are presented in Figures 2 and 3.

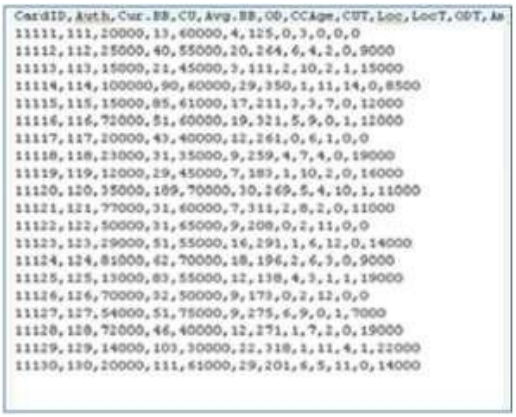

Fig. 4 – Dataset used for experiments

As seen in figure 2, the dataset contains credit card transaction details for a number of instances. This information is used by GA proposed in this paper. The GA makes use of the components as described in figure 3 so as to detect fraudulent transactions. The results of the detection are appeared in figure 3.

this paper help in understanding the fraud transactions and they can be used to prepare the system further so as to make new rules and achieve higher precision of fraud detection.

7. CONCLUSIONS

In this paper, we studied the problem of credit card fraud in E-commerce applications. We explored different approaches to solve the problem. The knowledge of different approaches can improve the scope of protecting E-commerce applications. At last, we implemented a genetic calculation for credit card fraud detection. As the adversaries change their means of assault every time, it is critical to have consistent vigil on the methods they use and update the techniques as needs be. In this paper manufacture a prototype application in Java stage so as to demonstrate the evidence of concept. The application uses a genetic calculation to detect credit card fraud. Information mining and other techniques are available to solve this problem. However, we preferred GA as it is efficient in detecting credit card fraud. The experimental results reveal that the proposed application is useful and can be used in real-world systems.

8. REFERENCES

Tej Paul Bhatla, Vikram Prabhu & Amit Dua "Understanding Credit Card Frauds," 2003.

Linda Delamaire, Hussein Abdou, John Pointon, "Credit card fraud and detection techniques: a review," *Banks and Bank Systems,* pp. 57-68, 2009.

Barry Masuda, "Credit Card Fraud Prevention: A Successful Retail Strategy," *crime prevention,* Vol. 6, 1986.

Ezawa.K. & Norton.S,"Constructing Bayesian Networks to Predict Uncollectible Telecommunications Accounts," IEEE *Expert,* October;45-51, 1996.

Peter J. Bentley, Jungwon Kim, Gil- Ho Jung and Jong-Uk Choi, "Fuzzy Darwinian Detection of Credit Card Fraud," *In the 14th Annual Fall Symposium of the Korean Information Processing Society,* 14th October 2000.

Amlan Kundu, S. Sural, A.K. Majumdar, "Two-Stage Credit Card Fraud Detection Using Sequence Alignment," *Lecture Notes in Computer Science, Springer Verlag, Proceedings of the International Conference on Information Systems Security*, Vol. 4332/2006, pp.260- 275, 2006.

Amlan Kundu, Suvasini Panigrahi, Shamik Sural and Arun K.Majumdar, "BLAST-SSAHA Hybridization for Credit Card Fraud Detection," *IEEE Transactions On Dependable And Secure Computing*, vol. 6, Issue no. 4, pp.309-315, October-December 2009.

Philip K. Chan ,Wei Fan, Andreas L. Prodromidis, Salvatore J. Stolfo, "Distributed Data Mining in Credit Card Fraud Detection,"
IEEE Intelligent Systems ISSN, Vol. 14 , Issue No. 6, Pages: 67 – 74, November 1999.

C. Phua, V. Lee, K. Smith, R. Gayler, "A Comprehensive Survey of Data Mining-based Fraud Detection Research," *Artificial Intelligence Review*, 2005.

Ray-I Chang, Liang-Bin Lai, Wen- De Su, Jen-Chieh Wang, Jen-Shiang Kouh, "Intrusion Detection by Backpropagation Neural Networks with Sample-Query and Attribute-Query," *Research IndiaPublications*, pp.6-10, November 26, 2006.

Ghosh, D.L. Reilly, "Credit Card Fraud Detection with a Neural-Network," *Proceedings of the International Conference on System Science*, pp.621-630, 1994.

. Brause, T. Langsdorf, M. Hepp, “Neural Data Mining for Credit Card Fraud Detection, *“International Conference on Tool with Artificial Intelligence*, pp.103-106,1999.

an, W. Miller, M.Stolfo, S.Lee & P Chan, “Using Artificial Anomalies to Detect Unknown and Known Network Intrusions,” anoel Fernando Alonso Gadi, Xidi Wang, Alair Pereira do Lago, “Credit Card Fraud Detection with Artificial Immune System,” *Lecture Notes in Computer Science*, Vol. 5132/2008, pp.119-131, 2008.

Aleskerov, B. Freisleben, B. Rao, “CARDWATCH: A Neural Network Based AUTHOR PROFILE Database Mining System for Credit Card Fraud Detection,” *Proceedings of IEEE/IAFE Conference on Computational Intelligence for Financial Engineering (CIFEr),* pp.220-226, 1997.

Benson Edwin Raj and A. Annie Portia, “Analysis on Credit Card Fraud Detection R Methods”, ICCCET, 2001, p1-5

Mehdi, S. Zair, A. Anou and M. Bensebti,” A Bayesian Networks in Intrusion Detection Systems,” *International Journal of Computational Intelligence Research*, Issue
No. 1, pp.0973-1873 Vol. 3, 2007. F

am, Bacchus, “Learning bayesian belief networks: An approach basedon the MDL principle,” *Computational Intelligence*, Vol.
10, Issue No. 3, pp.269–293, August 1994. M

am Maes, Karl Tuyls, Bram Vanschoenwinkel, Bernard Manderick, “Credit card fraud detection using Bayesian and neural networks,” *Interactive image-guided neurosurgery*, pp.261-270, 1993.4 E