



Survey Paper On Different Types Of Techniques used for Credit Card Fraud Detection

1 Nissy sarah Thomas; ²Stephie sara philipose; ³Sam G.Benjamin; ⁴Dr.B.Radhakrishnan

1 CSE Department, BMCE
APJ Abdul Kalam Technological
University, Sasthamcotta, Kerala, India.

2 CSE Department, BMCE
APJ Abdul Kalam Technological
University, Sasthamcotta, Kerala, India.

3 CSE Department, BMCE
APJ Abdul Kalam Technological
University, Sasthamcotta, Kerala, India.

4 CSE Department, BMCE

ABSTRACT

Credit Card fraud detection is considered as one of the major challenge that the digital world witnessed today. During this pandemic situation the use of credit card transactions has been increased tremendously. Hence fraud transactions also. The credit card fraudulent becomes a major security threat faced by the financial industry. Many modern techniques like genetic algorithm, machine learning, data mining, fuzzy logic etc has been used vividly used for detecting such fraud transactions. A clear analyzing and sound understanding about all these techniques will give an efficient credit card fraud detection system. This paper aims to presents survey of various techniques used in credit card fraud detection and evaluates each approach based on defined design criteria.

Key words: Machine Learning, Prediction, Classification Technique, Accuracy, Hidden Markovt, Logistic Regression

I. INTRODUCTION

Credit card is a type of payment card which is issued to users by financial firms. Cardholder can buy goods and services depending on the card holder's promise to pay for those goods and services. Due to the large scale popularity of the internet transactions for online shopping and globalization leads to the rise in credit card transactions also fraudulent associated with it also. Four types of credit card fraud are shown in below.

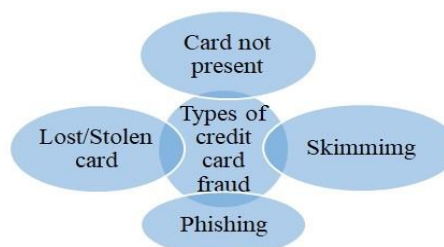


Figure 1:Types of credit card fraud

In Card not present, fraud customer does not physically present the card to the merchant during the fraudulent transaction. In Skimming crooks, use a small device to steal

credit card information in a card or from debit card transaction. When a credit or debit card is swiped through a skimmer, the device captures and stores all the details stored in the card's magnetic stripe. In phishing, Scammers might use a range of schemes to lure users into giving them their card info through tricks corresponding to websites simulation to be of a bank or payment system. When card is steal or lost, there are chances for a thief that he make unauthorized transaction before cardholder block the card.

II. BACKGROUND KNOWLEDGE

Classification, clustering, prediction, outlier detection, regression and visualization are the main six data mining techniques. These mining techniques are explained below:

A. CLASSIFICATION:

Classification is the problem of identifying to which of a set of categories, a new observation belongs to on the basis of a training set of data containing observations and whose categories membership is known. The application includes detection of credit card fraud, healthcare fraud, automobile insurance, corporate fraud, etc. where classification become very useful to detect fraud. **B..REGRESSION:**

Regression in machine learning consists of mathematical methods that allow data scientists to predict a continuous trend based on the value of one or more predictor variables. Regression is suitable for statistical method. Regression shows the relationship between more than one dependent and independent variable

C. PREDICTION:

Prediction is used to predict the continuous value. Based on the historic data, it make patterns, estimate the numeric, and ordered value for future. Predicted values are continuous value rather than discrete value.

D. CLUSTERING:

clustering performs by grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups. It is a variant of unsupervised While dissimilar objects are not become the part of that cluster; it transfer to another cluster

E. OUTLIER DETECTION:

Outliers are extreme values that deviate from other regular observations of normal data, they may indicate a variability in a measurement, experimental errors or a novelty Outlier detection means measure that distance between the data points and that outlier object

III. LITERATURE STUDY

Srivastava et al. [1], proposed a credit card fraud detection model based on Hidden Markov Model.HMM models the flow of operations in credit card transactions. HMM is a double embedded stochastic process. It includes a number of states. It is most efficient than a markov model. An HMM consists of number of states, number of distinct observation symbols per state, the state transition probability matrix, the observation symbol probability matrix, the initial state probability vector, the observation sequence. HMM does not accept incoming transaction with high probability, then it is recognized as fraudulent transaction. In credit card fraud detection HMM is initially trained with normal behaviour of a card holder. The fraud detection system checks each incoming transactions and get card details and value of purchase in order to verify whether the transaction is normal or fraudulent. If the transaction is fraud, then FDS raises an alarm and the issuing bank declines the transaction

Ms.Peter J. BentleyMs.Jung won Kim, Ms.Gil-Ho Jung and Ms.Jong-Uk Choi [2] proposed a credit card fraud detection system based on an evolutionary-fuzzy system capable of classifying suspicious and non- suspicious credit card transactions. The Fuzzy Darwinian Detection system uses genetic programming to evolve fuzzy logic rules capable of classifying credit card transactions into suspicious and non-suspicious classes. The system comprises a Genetic Programming (GP) search algorithm and a fuzzy expert system. The system first clusters the data into three groups namely low, medium and high. The GP ,the genotypes and phenotype of the GP System consist of rules which match the incoming sequence with the past sequence. Genetic Programming is used to evolve a series of variable-length fuzzy rules which characterize the differences between classes of data held in a database.The system is being developed with the specific aim of insurance-fraud

detection which involves the challenging task of classifying data into the categories into either safe or fraud. When the customer's payment is not overdue or the number of overdue payments is less than three months, the transaction is considered as non suspicious, otherwise it is considered as suspicious. The Fuzzy Darwinian detects suspicious and non -suspicious data and it easily detects stolen credit card Frauds.

Zeager MF [3] proposed a credit card fraud detection system using logistic regression. This model is an artificial intelligence based system. AI-based system detect fraud, needs a database to train the system. The data in real world are dirty and have missing values, noisy data, and outliers. Hence this model employed the mean-based method and clustering-based method for data cleaning. Cross validation technique is used for training the classifier. This ensures that the entire database is used for both the training data set and testing data set. Based on the accuracy, sensitivity, and error rate metrics the proposed classifier is evaluated. The proposed logistic regression based classifier is compared with other existed well known classifier, Knearest neighbors classifier and the voting classifier. The logistic regression-based classifier generates the best results among others.

Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang at [4] proposed a convolutional neural network(CNN) based approach to find fraudulent transactions in credit card. CNN is a kind of feed forward network. Generally CNN is designed for video recognition, image recognition. As part of deep learning input features are transformed into feature matrices and then converted into images. For finding more complex fraud patterns and to improve classification accuracy, a new feature trading entropy is proposed. To relieve the problem of the imbalanced dataset, they used cost based sampling method to generate different number of synthetic frauds to train the model. They applied CNN model because it is suitable for training large size of data and CNN has mechanism to avoid over-fitting.

Sahin and Duman [5] has built a model using decision tree and Support Vector Machine algorithm. They build fraud detecting model for the improvement of the financial transaction systems in an effective way. This work

demonstrates the uses of applying the data mining approaches including decision trees and SVMs to the credit card fraud detection problem with the real data set. In this study, the performance of classifier models built by decision tree methods C5.0, C&RT and CHAID and a number of different SVM methods are compared. When the performances of the models are compared with respect to accuracy, it is observed that as the number of the training data increases, this over fitting behavior becomes less remarkable and the performances of the SVM based models become comparable to decision tree based models. But the number of frauds caught by SVM models is less than the decision tree models, especially C&RT model. Though C5.0 model is the champion over the other models with respect to accuracy for each sample, C&RT model catches the largest number of frauds. So the C&RT and C5.0 models are choose as the final methods to build the prediction model. As a future work, other data mining algorithms such as different versions of Artificial Neural Networks (ANN) and logistic regression will be used to build new classification models on the same real world dataset and the performance of the new models will be compared with the performance of the models given in this paper.

C. Sudha ,T. Nirmal Raj [6] proposed a credit card fraud detection system with K Nearest Neighbour algorithm. One of the best classifier algorithms that have been used in credit card fraud detection is knearest neighbour algorithm. It is a supervised learning algorithm where the result of new instance query is classified based on majority of K-Nearest Neighbour category. A KNN algorithm is an evolutionary search and optimization technique that Mimics natural evolution to find the best solution to a problem. Here they take IP address, spending habits, cardholder name as parameters. This system classify any incoming transaction by calculating the nearest point to the new incoming transaction. Then if the nearest neighbour is fraudulent, then the transaction indicates fraud. The value of K is used as a small and odd to break the ties They used OTP for security of the model, This method proves accurate in detecting fraudulent transactions and minimizing the number of false alerts.

Ghosh and Reilly [7] proposed a method based on Artificial Neural Network(ANN).They used data from a credit card issuer. It was a three-layer feed forward neural network . ANN is one of the most powerful classifier to find out the hidden patterns among different attributes. The working of ANN is similar to human brain. It consist of various layers. The first layer is called input layer and the last one is called output layer. It may or may not have hidden layer. Each neuron has different layers and each has a weighted edge. Neural network was trained on a large sample of labeled credit card account transactions and tested on a holdout data set that consisted of all account activity over a subsequent two-month period of time.

Wen-Fang YU and Na Wang [8] proposed a method based on Outlier mining, Outlier detection mining and Distance sum algorithms are used to accurately predict fraudulent transaction. They used dataset of credit card transaction certain commercial bank. Outlier mining is a field of data mining which is basically used in monetary and internet fields ,is based on detecting items that are deviated from the main system i.e. the transactions that are fraud. The attributes used are customer's behavior and with those values of attributes they calculated that distance between the observed value of that attribute and its predetermined value.one of the major advantage of this method is the

probability of fraud transactions can be predicted soon after credit card transactions by the banks.

Ishu Trivedi, Monika, Mrigya Mridushi [9] proposed a credit card fraud detection system using genetic algorithms.Genetic algorithms consist of approaches for calculating optimal solution for the problem and implicitly producing the result of the fraudulent transaction. Genetic algorithm is a heuristic approach used to solve high complexity computational problems.The operators of Genetic Algorithm are: Selection , Mutation ,Crossover.The algorithm reduces the number of false alerts. The model will detects fraud transaction in short span of time.

Zhang Yongbin et al[10] proposed a behaviour based credit card model. This model analyses the behaviour pattern of cardholder. This model will check the historical pattern of the cardholder. If any deviation is found then it is considered to be fraud. They uses transaction record of a single credit card user to build the model. In this model Self organizing map method is used to detect the outliers from the normal ones. The main focus is on money withdrawing transactions from bank with credit card.SOM is an unsupervised learning algorithm, the outputs of SOM are used to judge if the current ongoing transaction is legal or suspected according to the majoritarian strategy adopted.

TABLE I. A COMPARATIVE STUDY OF VARIOUS ALGORITHMS IN LITERATURE REVIEW

Title	First author	Journal	Method	advantages	Disadvantages
Credit Card Fraud Detection Using Hidden Markov Model [1]	Avhinav Srivastava	IEEE Dep & Sec Comp. 2002	Hidden Marcov Model	Can handle large volume of data	High false rate

Fuzzy Darwinian Detection of Credit Card Fraud[2]	Peter j Bentley	Symposium of the Korean Information Processing Society	Fuzzy logic and genetic algorithm	good accuracy .Low false alarm	Cannot be used in online transaction. Processing speed is low, high cost
Adversarial Learning in Credit Card Fraud Detection [3]	Frances Zeager	IEEE - Conference 2017	Logistic regression	Data imbalance is solved	classifier retraining cost was not addressed

Credit Card Fraud Detection using Convolutional Neural Network [4]	Kang Fu	Springer,2016	Convolutional Neural Network	Less time for training	Take large time for detection
CREDIT CARD FRAUD DETECTION USING RANDOM FOREST ALGORITHM[5]	Sahin and Duman	International Conference on Computing and Communication Technologies ICCCT 2019	Random Forest	Can run on large database. Accurate classifier	Overfit for noisy data
CREDIT CARD FRAUD DETECTION IN INTERNET USING KNEAREST NEIGHBOR ALGORITHM[6]	C. Sudha	IPASJ International Journal of Computer Science (IJCS)2017	K nearest algorithm	Best performance	Accuracy based on neighbours distance fraud cannot be detected at the time of transaction.
Credit Card Fraud Detection using Artificial Neural Network [7]	Ghosh and Reilly	IEEE - Conference 2019	Artificial neural network	Real time detection	More complex More prediction time
Research on Credit Card Fraud Detection Model Based on Distance Sum[8]	Wen-Fang YU and Na Wang	IEEE Conference Artificial Intelligence	Outlier detection	Works fast and well on online large datasets	Cannot detect fraud at the time of transaction
Credit Card Fraud Detection [9]	Ishu Trivedi, Monika	International Journal of Advanced Research in Computer and Communication Engineering	Genetic algorithm	Inexpensive and fast in detection	Requires extensive tool knowledge to set up and operate and difficult to understand
Behavior-based credit card fraud detecting model[10]	Zhang Yongbin	IEEE.	Self-Organising Maps	Easy to train	Time for detecting fraud in the real-time. difficult to obtain perfect mapping

IV.CONCLUSION

This paper presents a detailed survey of various techniques used for credit card fraud detection. Every method has its own merits and demerits, one of these or hybrid techniques can be used for detection. New features can be included to improve the accuracy of the model. To avoid data imbalance problem various sampling techniques can be established.

REFERENCE

- [1]Srivastava A, Kundu A, Sural S, Majumdar A. Credit card fraud detection using hidden Markov model. IEEE Transactions on dependable and secure computing. 2008 Feb 8;5(1):37-48.
- [2] Bentley PJ, Kim JW, Jung GH, Choi JU. Fuzzy darwinian detection of credit card fraud. InProceedings of the Korea Information Processing Society Conference 2000 (pp. 277-280). Korea Information Processing Society.
- [3] Zeager MF, Sridhar A, Fogal N, Adams S, Brown DE, Beling PA. Adversarial learning in credit card fraud detection. In2017 Systems and Information Engineering Design Symposium (SIEDS) 2017 Apr 28 (pp. 112-116). IEEE.
- [4] Fu K, Cheng D, Tu Y, Zhang L. Credit card fraud detection using convolutional neural networks. InInternational Conference on Neural Information Processing 2016 Oct 16 (pp. 483-490). Springer, Cham.
- [5] Y. Sachin, E. Duman, "Detecting Credit Card Fraud by Decision Tree and Support Vector Machine", In Proceedings of the international multi Conference of Engineers and Computer Scientists, Hong Kong, 2011, pp..
- [7] Ghosh S, Reilly DL. Credit card fraud detection with a neural-network. InSystem Sciences, 1994. Proceedings of the Twenty-Seventh Hawaii International Conference on 1994 Jan 4 (Vol. 3, pp. 621-630). IEEE.
- [8] Yu WF, Wang N. Research on credit card fraud detection model based on distance sum. In2009 International Joint Conference on Artificial Intelligence 2009 Apr 25 (pp. 353-356). IEEE.
- [9] Trivedi I, Monika MM. Credit card fraud detection. International Journal of Advanced Research in Computer and Communication Engineering. 2016 Jan;5(1).
- [10]Zhang Y, You F, Liu H. Behavior-based credit card fraud detecting model. In2009 Fifth International Joint conference on INC, IMS and IDC 2009 Aug 25 (pp. 855-858). IEEE.

