Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IMPACT FACTOR: 6.199

IJCSMC, Vol. 8, Issue. 8, August 2019, pg.43 – 49

Credit Card Fraud Detection Techniques: A Review

Sonal Mehndiratta
Mtech Scholar
sonalmehndiratta02@gmail.com
Guru Nanak Institute of Technology
Mullana, Ambala

Mr. Kamal Gupta
Hod and Assistant professor
Kamalgupta123@gmail.com
Guru Nanak Institute of Technology
Mullana, Ambala

Abstract: The prediction analysis is the approach which can predict future possibilities on the current data. When the physical-card based purchasing technique is applied, the card is given by the cardholder to the merchant so that a successful payment method can be performed. The fraudulent transactions are conducted by the attacker by stealing the credit card. When the loss of the card is not noticed by the cardholder, a huge loss can be faced by the credit card company. A very little amount of information is required by the attacker for conducting any fraudulent transaction in online transactions. In this research work, various credit card fraud detection techniques are reviewed in terms of certain parameters.

Keywords: Machine learning, Classification, Credit card fraud Detection

Introduction

Data Mining is the process which is applied to extract relevant data from the rough data. The consistent patterns and systematic relationships amongst variables are searched through the analytic process. The patterns which were detected previously from the input data are applied to new subsets of data. The useful information is analyzed and extracted on the basis of certain algorithms. This extracted information then helps to discover the hidden patterns and relationships from data automatically [1]. The prediction analysis is most useful type of data which is performed today. To perform the prediction analysis the patterns needs to generate from the dataset with the machine learning. The prediction analysis can be done by gathering historical information to generate future trends. So, the knowledge of what has happened previously is used to provide the best valuation of what will happen in future with predictive analysis [2]. The predication analysis models are designed according to application type. The model is trained using the sample data that includes known attributes. The new data can be analyzed and its behavior can be determined using this trained model. Credit card fraud detection is one of the applications of prediction analysis. In credit card fraud detection, the fraud transactions are predicted based on the historical information of credit card transactions. The historical information of card transactions will be the training data for the fraud transaction prediction [3]. The credit cards are being used very commonly today for buying several goods and accessing various services in our daily lives. When the physical-card based purchasing technique is applied, the card is given by the cardholder to the merchant so that a successful payment method can be performed. The fraudulent transactions are conducted by the attacker by stealing the credit card. When the loss of the card is not noticed by the cardholder, the huge loss can be faced by the credit card company [4]. A very little amount of information is required by the attacker for conducting any fraudulent transaction in online transactions. For buying products and services online, the Internet or telephone devices are used. In some cases, the pattern in which transactions are done by the user is the only way through which it is possible to know that the card is stolen. A fraud detection method needs to be applied to reduce the rate of successful credit card frauds. The existing purchase data of the particular cardholder is the basis on which this fraud detection method is proposed. One of the biggest threats being faced by business organizations today is credit card fraud. The approaches that result in causing fraud need to be perceived initially so that they can be handled in an effective manner. For committing the fraud, a variety of methods are used by credit card fraudsters. When someone else's credit card is used for personal reasons and the owner of the card does not have any knowledge about it, the credit card fraud is outlined [5]. The person who is conducting the fraud will never aim to contact the owner of the card or repay the losses to the actual user. There are two significant categorizations of the credit card fraud detection techniques. They are fraud analysis which is also called misuse detection and user behavior analysis which is also called anomaly detection. The anomaly detection techniques can detect the anomalies based on the previous information available. The anomaly detection techniques are broadly classified into two categories. The supervised learning is the efficient approach for the credit card fraud detection. In the supervised learning approach the labels are assigned to the training set for the classification [6]. The classification approach will assign the designer labels to test set. Basically, this approach focuses on training the machine using a properly labeled data. Once the training process is complete, new set of examples are provided to the machine and a correct outcome is achieved from the labeled data. There are two broad categorizations of supervised learning algorithms which are classification and regression. In case when the output variable is in the form of a specific category, the classification method is considered. Whereas, when the output is achieved as a real value like dollars or weight, a regression method is followed. The unsupervised learning approach is less efficient as compared to supervised learning approach. In the unsupervised learning approach the labels are not assigned to training set [7]. The technique needs to apply which assign label to the training set to get final result of classification. Grouping the unsorted type of data on the basis of the patterns, similarities and differences without performing any prior training on it, is the main function of unsupervised learning approach. In the unlabeled data, the hidden structure is to be identified by the machine on its own. Clustering and association are the two categories in which the unsupervised algorithms are classified. The inherent groupings within the data are discovered through clustering. Grouping the customers based on their purchasing behavior is one of the examples of clustering. Discovering the rules that describe large portions of data is done through association [8]. Different techniques have been applied to detect the frauds that occur in credit card transactions. These techniques are explained below:

a. Artificial Neural Network: A set of interlinked nodes that are designed for imitating the working of a human brain is known as an artificial neural network (ANN). A weighted link is assigned to all the other nodes that are present in the adjacent layers of each node.

b. Genetic Algorithm (GA): The genetic algorithms were introduced inspiring from natural evolution. Chromosomes are the binary strings that are used to represent the populations of candidate solutions. It is based on the concept that the chances of survival and reproduction are higher for the chromosomes with higher quality i.e. having better fitness value.

c. Hidden Markov Model (HMM): A double embedded stochastic process using which highly complicated stochastic processes can be generated is known as a hidden Markov model [9]. A Markov process that has unobserved states is assumed to be available within the underlying system. The only unknown parameters are the definite transition of the states within the simpler Markov models.

d. KNN Classifier: KNN is the non-parametric algorithm used in case of classification and regression. In classification and regression, the input is consisting of K-nearest training examples in the feature space and on the other hand, the output depends upon whether KNN belongs to regression category or classification category.

e. Naïve Bayes: This algorithm implements the Bayesian rule on categorical data for performing classification on it. In comparison to other classification approaches, the performance of the Naïve Bayes algorithm is known to be better and very simple.

Literature Review

Kuldeep Randhawa et al. [10] proposed a technique using machine learning to detect credit card fraud detection. Initially, standard models were used after that hybrid models came into picture which made use of AdaBoost and majority voting methods. Publically available data set had been used to evaluate the model efficiency and another data set used from the financial institution and analyzed the fraud. Then the noise was added to the data sample through which the robustness of the algorithms could be measured. The experiments were conducted on the basis of the theoretical results which show that the majority of voting methods achieve good accuracy rates in order to detect the fraud in the credit cards. For further evaluation of the hybrid models noise of about 10% and 30% has been added to the sample data. Several voting methods have achieved a good score of 0.942 for 30% added noise. Thus, it was concluded that the voting method showed much stable performance in the presence of noise.

Abhimanyu Roy et al. [11] proposed deep learning topologies for the detection of fraud in online money transaction. This approach is derived from the artificial neural network with in-built time and memory components like long term short term memory and several other parameters. According to the efficiency of these components in fraud detection, almost 80 million online transactions through credit card have been pre-labeled as fraudulent and legal. They have used high performance distributed cloud computing environment. The study proposed by the researchers provides an effective guide to the sensitivity analysis of the proposed parameters as per the performance of the fraud detection. The researchers also proposed a framework for the parameter tuning of Deep Learning topologies for the detection of fraud. This enables the financial institution to decrease the losses by avoiding fraudulent activities.

Shiyang Xuan et al. [12] used two types of random forests which train the behavior features of normal and abnormal transactions. The researcher compares these two random forests which are differentiated on the basis of their classifiers, performance on the detection of credit card fraud. The data used is of an e-commerce company of China which is utilized to analyze the performance of these two types of random forests model. In this paper, the author has used B2C dataset for the identification and detection of fraud from the credit cards. Therefore, the researcher concluded from the result that the proposed random forests provide good results on small dataset but there are still some problems like imbalanced data which makes it less effective than any other dataset.

Zahra Kazemi et al. [13] proposed Deep autoencoder which is used to extract the best characteristics of the information from the credit card transaction. This will further add softmax software to resolve the class labels issues. An overcomplete autoencoder is used to map the data into a high dimensional space and a sparse model was used in a descriptive manner which provides benefits for the classification of a type of fraud. Deep learning is one of the most motivated and powerful techniques being employed for the detection of fraud in the credit card. These types of networks have a complex distribution of data which is very difficult to recognize. Deep autoencoder has been used in some stages to extract the best features of the data and for the classification purposes. Also, higher accuracy and low variance are achieved within these networks.

John O. Awoyemi et al. [14] proposed an investigation through which the performances of several algorithms were evaluated when they were applied on credit card fraud data that is highly skewed. The European cardholders' 284,807 transactions were used as a source to generate the dataset of credit card transactions. On the skewed data, a hybrid approach of under-sampling and oversampling is performed. On raw and preprocessed data, there are three different techniques applied in Python. Based on certain parameters like precision, sensitivity, accuracy, balanced classification rate and so on, the performances of these techniques are evaluated. It is seen through the achieved results that in comparison to naïve Bayes and logistic regression approaches, the performance of k-NN is better.

Sharmistha Dutta et al. [15] presented a study on the commonly found crime within the credit card applications. There are certain issues faced when the existing non-data mining approaches are applied to avoid identity theft. A novel data mining layer of defense is proposed for solving these issues. For detecting the frauds within various applications, two algorithms named Communal Detection and Spike Detection which generate novel layer. There is a large moving window, higher numbers of attributes and numbers of link types available which can be searched by CD and SD algorithms. Thus, results can be generated by the system by consuming a huge amount of time. Since the attackers do not get time to modify their behaviors with respect to the algorithms being deployed in real time, there is no true evaluation achieved even after a regular update of the algorithms. Therefore, it is not possible to properly demonstrate the concept of adaptability. These issues can be resolved by making certain enhancements in the proposed algorithm in future work.

Krishna Modi et al. [16] investigated several techniques that were used for detecting the fraudulent transactions and provided a comparative study amongst them. The fraudulent transactions can be detected by utilizing either one of these or integrating any of these methods. The model can possibly be trained in a more accurate manner by adding new features. Several data mining techniques are being used by bank and credit card companies for detecting fraud behaviors. The normal usage pattern of clients depending upon their past activities can be identified by applying any of these methods. Therefore, a comparative analysis is made here by studying different fraud detection techniques proposed over the years.

Dastgir Pojee et al. [17] proposed a novel mechanism using which the payment of invoice or bill is initiated. This approach is named as 'NoCash' mobile application which is mainly used by the merchants through which the payment facility of clients can be eased. There is no need for NFC-Enabled Point of Sales (PoS) Machines in this approach and only the mobile phones are required. Minimizing the burden of clients for bringing cards when outside, by providing easy payment transferring mechanisms is the only aim for which this system is designed. The client's experience of shopping is improved when NoCash application that includes many features is applied on the basis of the increase in a number of NFC-based mobiles. To provide benefits to merchants, the fraud activities are minimized using this proposed application. The application clients can be related to the expense history and minimize any unwanted costs using this proposed method.

Dilip Singh Sisodia et al. [18] presented the evaluation of the performance of several sampling techniques on the classifier when they are applied on credit card fraud data set with the class imbalance. The principal component analysis (PCA) is applied to real data as well as the variables time, amount and class to achieve 28 principal components that are included within the data. There are ten thousand, fifteen thousand and twenty thousand instances available respectively within the three datasets. This approach applied five over-sampling and four under-sampling approaches. Further, on the data, few cost-sensitive and ensemble classifiers are applied.

Luis Vergara et al. [19] proposed an improvement in the performance of credit card fraud detection by developing various methods that are based on signal processing. A variant of the traditional iterative amplitude adjusted Fourier transform (IAAFT) and the iterative surrogate signals on graph algorithms (ISSG) are present within the proposed methods. Improving the training of detectors is the major aim of this approach. The surrogate samples are generated from original fraud samples in this mechanism. The variance of the estimate is reduced here such that the training of detectors can be improved. Due to the presence of various issues and the constant change of patterns present in the data stream it is important to provide a reliable augmentation of the target scarce population of frauds. The real data was used in this experiment to demonstrate the capabilities of proposed methods such that the performance of detection can be improved. The ROC curves and KPIs which are commonly used in financial business were used in this research to measure the capabilities.

Table of Comparison

Author's Name	**Technique**	**Advantages/ Features**	**Disadvantages/ Improvements**
Kuldeep Randhawa, Chu Kiong Loo, Manjeevan Seera, Chee Peng Lim and Asoke K. Nandi	Machine learning to detect credit card fraud detection.	Majority of voting methods achieve good accuracy rates in order to detect the fraud in the credit cards.	The precision value achieved is less as compared to other algorithms.
A. Roy and J. Sun and R. Mahoney and L. Alonzi and S. Adams and P. Beling	Deep learning topologies for the detection of fraud in online money transaction.	Proposed model outperformed and prevented the frauds in any online transaction through credit cards	There is a need to improve the accuracy of the proposed algorithm.
Guanjun Liu, Zhenchuan Li, Lutao Zheng, Shuo Wang and Changjun Jiang Shiyang Xuan	The B2C dataset for the identification and detection of fraud from the credit cards.	Proposed random forests provide good results on the small dataset	Problems like imbalanced data make it less effective than any other dataset.
Zarrabi, H. Kazemi	Deep autoencoder which is used to extract the best characteristics of the information from the credit card transaction.	The classification is performed on the best-extracted features. Due to which it gains high accuracy, the low variance is noticeable.	The less number of variation are used to examine the results of the proposed approach
John O., Adebayo O. Adetunmbi, and Samuel A. Oluwadaren Awoyemi	The performances of several algorithms were evaluated when they were applied on credit card fraud data that is highly skewed.	In comparison to naïve Bayes and logistic regression approaches, the performance of k-NN is better.	Limited parameters are used to test the performance level.
S. Dutta, A. K. Gupta and N. Narayan	Novel data mining layer of defense is proposed using two algorithms named Communal Detection and Spike Detection.	There is a large moving window, higher number of attributes and number of link types available which can be searched by CD and SD algorithms.	The evaluation of the results is not done properly and also presided information is given about result analysis
K. Modi and R. Dayma	Several methods are integrated to provide a secure mechanism.	The normal usage pattern of clients depending upon their past activities is identified by applying any of these methods.	The proposed algorithm achieves high performance in terms of execution time by accuracy factor get compromised.

D. Pojee, S. Zulphekari, F. Rarh, and V. Shah	'NoCash' mobile application is proposed.	The fraud activities are minimized using this proposed application	The expense history and any unwanted costs need to be minimized.
D. S. Sisodia, N. K. Reddy and S. Bhandari	The principal component analysis (PCA) is applied to real data to propose a novel approach.	The performance of various methods was evaluated using certain performance metrics which showed the proposed approach's efficiency against others.	The precision value and execution time are not as per the demand.
L. Vergara, A. Salazar, J. Belda, G. Safont, S. Moral and S. Iglesias	A variant of the traditional iterative amplitude adjusted Fourier transform (IAAFT) and the iterative surrogate signals on graph algorithms (ISSG) are proposed.	The real data was used to evaluate the performances of the proposed method which showed the efficiency of the proposed approach.	Precision was low as compared to other algorithms.

Conclusion

The fraud transaction detection is the major issue of prediction due to a frequent and large number of transactions. The fraud transaction prediction has the two phases which are feature extraction and classification. In the first phase, the feature extraction technique is applied and in the second phase, classification is applied for the fraud transaction detection. In this review paper various techniques of credit card fraud detection are reviewed. In future hybrid approach will be designed for the credit card fraud detection.

References

[1] S.B.E. and Portia, A.A., Raj, "Analysis on credit card fraud detection methods, "*International Conference on Computer, Communication and Electrical Technology (ICCCET)*, pp. 152-156, 2015.

[2] Rajni, Bhupesh Gour, and Surendra Dubey Jain, "A hybrid approach for credit card fraud detection using rough set and decision tree technique," *International Journal of Computer Applications*, vol. 139, no. 10, pp. 1-6, 2016.

[3] Agrawal A.N Dermala N., "Credit card fraud detection using SVM and Reduction of false alarms," *International Journal of Innovations in Engineering and Technology (IJIET)*, vol. 7, no. 2, pp. 176-182, 2016.

[4] Phua C., Lee V., Smith, Gayler K.R., "A comprehensive survey of data mining-based fraud detection research", arXiv preprint arXiv:1009.6119, 2010.

[5] Stojanovic A., Aouada D., Ottersten B Bahnsen A.C., "Cost-sensitive credit card fraud detection using Bayes minimum risk," in *12th International Conference on Machine Learning and Applications (ICMLA)*, pp. 333-338, 2013.

[6] Carneiro E.M., Dias L.A.V., Da Cunha A.M., Mialaret L.F.S., "Cluster analysis and artificial neural networks: A case study in credit card fraud detection", *in 12th International Conference on Information Technology-New Generations*, pp.122-126, 2015.

[7] S. Aghili and P. Zavarsky K. T. Hafiz, "The use of predictive analytics technology to detect credit card fraud in Canada," in *11th Iberian Conference on Information Systems and Technologies (CISTI)*, pp. 1-6, 2016.

[8] Bansal M Sonepat H.C.E., "Survey Paper on Credit Card Fraud Detection," *International Journal of Advanced Research in Computer Engineering & Technology*, vol. 3, no. 3, pp. 827-832, 2014.

[9] S., Tuyls, K., Vanschoenwinkel, B. and Manderick, "Credit card fraud detection using Bayesian and neural networks," in *Proceedings of the 1st international naiso congress on neuro-fuzzy technologies*, pp. 261-270, 2002.

[10] Kuldeep Randhawa, Chu Kiong Loo, Manjeevan Seera, Chee Peng Lim and Asoke K. Nandi, "Credit card fraud detection using AdaBoost and majority voting," *IEEE Access*, vol. 6, pp. 14277-14284, 2018.

[11] A. Roy and J. Sun and R. Mahoney and L. Alonzi and S. Adams and P. Beling, "Deep learning detecting fraud in credit card transactions," in *Systems and Information Engineering Design Symposium (SIEDS)*, pp. 129-134, 2018.

[12] Guanjun Liu, Zhenchuan Li, Lutao Zheng, Shuo Wang and Changjun Jiang Shiyang Xuan, "Random Forest for Credit Card Fraud Detection," in *IEEE 15th International Conference On Networking, Sensing and Control (ICNSC)*, pp.1-6, 2018.

[13] Zarrabi, H. Kazemi, "Using deep networks for fraud detection in the credit card transaction," *IEEE 4th International Conference In Knowledge-Based Engineering and Innovation (KBEI)*, pp. 0630-0633, 2017.

[14] John O., Adebayo O. Adetunmbi, and Samuel A. Oluwadaren Awoyemi, "Credit card fraud detection using machine learning techniques: A comparative analysis."*International Conference on Computing Networking and Informatics (ICCNI)*, pp. 1-9, 2017.

[15] S. Dutta, A. K. Gupta and N. Narayan, "Identity Crime Detection Using Data Mining, "*3rd International Conference on Computational Intelligence and Networks (CINE)*, Odisha, pp. 1-5, 2017.

[16] K. Modi and R. Dayma, "Review on fraud detection methods in credit card transactions, "*International Conference on Intelligent Computing and Control (I2C2)*, Coimbatore, pp. 1-5, 2017.

[17] D. Pojee, S. Zulphekari, F. Rarh, and V. Shah, "Secure and quick NFC payment with data mining and intelligent fraud detection, "*2nd International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, pp. 148-152, 2017.

[18] D. S. Sisodia, N. K. Reddy and S. Bhandari, "Performance evaluation of class balancing techniques for credit card fraud detection," *IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, Chennai, pp. 2747-2752, 2017.

[19] L. Vergara, A. Salazar, J. Belda, G. Safont, S. Moral and S. Iglesias, "Signal processing on graphs for improving automatic credit card fraud detection," *International Carnahan Conference on Security Technology (ICCST)*, Madrid, pp. 1-6, 2017.