



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING

Sonali Admane , Devanshi Gangrade , Nidhi Patil , Shradha Nazirkar , Madhuri Thorat

Department of Information Technology , AISSMS Institute of Information Technology , Pune – 01 , India

Abstract: Now a day's online payment gaining popularity because of easy and convenience use of ecommerce. It became very easy mode of payment. People choose online payment and e-shopping; because of time convenience, transport convenience, etc. As the result of huge amount of e-commerce use, there is a vast increment in credit card fraud also. Machine Learning has been successfully applied to finance databases to automate analysis of huge volumes of complex data. Machine Learning has also played a salient role in the detection of credit card fraud in online transactions. Fraud detection in credit card is a big problem, it becomes challenging due to two major reasons—first, the profiles of normal and fraudulent behaviors change frequently and secondly due to reason that credit card fraud data sets are highly skewed. This paper research and checks the performance of Random Forest on highly skewed credit card fraud data. Dataset of credit card transactions is sourced from European cardholders containing 1 lakh transactions. These techniques are applied on the raw and preprocessed data. The performance of the techniques is evaluated based on accuracy, sensitivity, and specificity, precision. . Keywords— Data Analysis, Fraud in credit card, Machine Learning, Security.

I. INTRODUCTION

Now a day's online payment gaining fashionability because of easy and convenience use of e-commerce, it came veritably easy mode of payment. People choose online payment and e-shopping; because of time convenience, transport convenience, etc. As the result of huge quantum of e-commerce use, there's a vast proliferation in credit card fraud also. Machine Learning has been successfully applied to finance databases to automate analysis of huge volumes of complex data. Machine Learning has also played a salient part in the discovery of credit card fraud in online deals. Fraud discovery in credit card is a big problem, it becomes grueling due to two major reasons – first, the biographies of normal and fraudulent behaviours change constantly and secondly due to reason that credit card fraud data sets are largely slanted. Credit card fraud is a growing concern with far reaching consequences in the government, commercial associations, finance assiduity, In Moment's world high reliance on internet technology has enjoyed increased credit card deals but credit card fraud had also accelerated as online and offline sale. As credit card deals come a wide mode of payment, focus has been given to recent computational methodologies to handle the credit card fraud problem. There are numerous fraud discovery results and software which help frauds in businesses similar as credit card, retail, e-commerce, insurance, and diligence. Machine Learning is one notable and popular styles used in working credit fraud discovery problem. It's insolvable to be sheer certain about the true intention and wickedness behind an operation or sale. In reality, to seek out possible attestations of fraud from the available data using fine algorithms is the stylish effective option. Fraud discovery in credit card is the truly the process of relating those deals that are fraudulent into two classes of legal class and fraud class deals, several ways are designed and enforced to break to credit card fraud discovery similar as inheritable algorithm, artificial neural network frequent item set mining, migrating catcalls optimization algorithm, relative analysis of decision tree and arbitrary timber is carried out. Credit card fraud discovery is a veritably popular but also a delicate problem to break. Originally, due to issue of having only a limited quantum of data, credit card makes it grueling to match a pattern for dataset. Secondly, there can be numerous entries in dataset with truncations of fraudsters which also will fit a pattern of licit geste. Also the problem has numerous constraints. Originally, data sets aren't fluently accessible for public and the results of inquiries are frequently hidden and cleaned, making the results inapproachable and due to this it's grueling to benchmarking for the models erected. Datasets in former inquiries with real data in the literature is nowhere mentioned. Secondly, the enhancement of styles is more delicate by the fact that the security concern imposes a limitation to change

of ideas and styles in fraud discovery, and especially in credit card fraud discovery. Incipiently, the data sets are continuously evolving and changing making the biographies of normal and fraudulent actions always different that's the legal sale in the history may be a fraud in present or vice versa. This paper evaluates two advanced machine literacy, Decision tree and arbitrary timbers and also a collative comparison is made to estimate that which model performed stylish. Credit card sale datasets are infrequently available, largely imbalanced and slanted. Optimal point (variables) selection for the models, suitable metric is most important part of mining to estimate performance of ways on slanted credit card fraud data. A number of challenges are associated with credit card discovery, videlicet fraudulent geste profile is dynamic, that's fraudulent deals tend to look like licit bones, Credit card fraud discovery performance is greatly affected by type of slice approach used, selection of variables and discovery fashion used.

2.1 LITERATURE SURVEY

There's colorful Fire Discovery fashion is proposed by different authors as follows.

Hardik Manek et al (1) executed and compared Logistic Retrogression and Autoencoder Neural Network for fraud discovery in banking deals. Results supported our perpetration, Logistic Retrogression model helped to attain delicacy of 58. Since also, the definitive thing of adding delicacy was satisfied through the Autoencoder Neural Network, with delicacy of 99.83. Hence by assaying the result, it's observed that Autoencoder Neural Network has easily outperformed the Logistic Retrogression model. The reason behind the superior performance of Autoencoder Neural Network is it has the capability of tone- learning through backpropagation. Autoencoder Neural Network uses the sigmoid function as one of the activation function in the retired subcaste which means that Logistic Retrogression is a subset of Autoencoder Neural Network. There are multiple layers in Autoencoder Neural Network in our perpetration there are 4 layers which perform the distinct function of garbling, calculating, and decoding. While Logistic Retrogression is considered as a single subcaste neural network which alone performs the complete function of double bracket, thus, performing in reduced delicacy. Accordingly, indeed though the neural network is computationally more precious it has the advantage of learning more complex and non-linear functions.

In (2) this document proposes a new relative measure of the comparison rules that nicely represents the gains and losses due to fraud discovery. A cost-sensitive system grounded on the minimal Bayes threat is presented using the proposed cost measure. Advancements of over 23 are attained by comparing this system and other rearmost- generation algorithms. The data set for this document is grounded on the real- life transactional data of a large European company and particular data in the data is kept nonpublic. The delicacy of an algorithm is about 50. The significance of this work was to find an algorithm and reduce the cost dimension. The result was 23 and the algorithm they plant was the minimum threat of Bayes.

In (3) Several ultramodern ways grounded on sequence alignment, machine literacy, artificial intelligence, inheritable programming, data mining, etc. They've been developed and are still being developed to descry fraudulent credit card deals. A solid and clear understanding of all these approaches is demanded, which will really lead to an effective credit card fraud discovery system. This document shows a check of different ways used in credit card fraud discovery mechanisms and the evaluation of each methodology grounded on certain design criteria. An analysis of credit card fraud discovery styles was performed. The check in this document was grounded solely on detecting the effectiveness and translucency of each system. The significance of this document was to conduct a check to compare different credit card fraud discovery algorithms to find the most applicable algorithm to break the problem.

An approach is proposed towards fraud discovery in banking deals in (4) using fuzzy clustering and neural network. In this approach, fraud discovery is performed in three phases. The first step is to authenticate the original stoner and corroborate card details. A vague half-clustering algorithm is performed after completion of this operation to descry the geste of normal stoner use grounded on oncentransactions. However, grounded on a neural network to determine whether it was actually a fraudulent sale or whether the medium applies, If a new sale at this stage turns out to be uncertain.

Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang at (5) proposed a convolutional grounded neural network approach (CNN) to find fraudulent deals. Convolutional neural network is a part of deep literacy and is a type of advanced neural network composed of further than one hidden subcaste. In this document, chancing more complex models of fraud and perfecting bracket delicacy, a new point of marketable entropy is proposed. In this document for the first time, CNN is used to descry fraud used to descry frauds.

Krishna Modi and Reshma Dayma (6) applied neural network and Convolutional Neural network is applied with SMOTE. Also converted features and added new point that gives better performance by increased number of TP and dropped number of FP. Comparison result shows that in CNN, performance of perfection is poor than NN because rate of licit deals detected as fraudulent one is further than neural network. On other hand in CNN, rate of detecting fraudulent sale is further than NN which improves performance of recall and F1 score. Results show that CNN with SMOTE and point metamorphosis overcome issue of perfection and outperforms NN in all terms. Limitation is fraudulent sale which geste is same as licit deals can n't be detected.

Paper "Credit Card Fraud Discovery using Deep Literacy grounded on Bus-Encoder and Confined Boltzmann Machine" (7), aims to 1) concentrate on fraud cases that can not be detected grounded on former history or supervised literacy, 2) produce a model of deep Bus-encoder and confined Boltzmann machine (RBM) that can reconstruct normal deals to find anomalies from normal patterns. The proposed deep literacy grounded on bus-encoder (AE) is an unsupervised literacy algorithm that applies backpropagation by setting the inputs equal to the labors. The RBM has two layers, the input subcaste (visible) and hidden subcaste. In this exploration, we use the Tensorflow library from Google to apply AE, RBM, and H2O by using deep literacy. The results show the mean squared error, root mean squared error, and area under wind.

The outlier discovery is another system used to descry both supervised and unsupervised literacy. Supervised outlier discovery system studies and classifies the outlier using the training dataset. Again, unsupervised outlier discovery is analogous to clustering data into multiple groups grounded on their attributes. N. Malini and Dr. M. Pushpa (8) mention that the outlier discovery system grounded on unsupervised literacy is preferred to descry credit card fraud over outlier supervised literacy, because

unsupervised literacy outlier doesn't bear previous information to marker data as fraudulent. So, it needs to be trained by using normal deals to distinguish between a legal or illegal sale.

Some credit card fraud sale datasets contain the problem of imbalance in datasets. Anusorn Charleonnann (9) mentions that the unbalance of datasets has numerous characteristics that crop during the bracket. He uses RUS, a data slice fashion, by trying to relieve the problem of class unbalance by editing the class distribution of training datasets. There are two major styles of conforming the imbalance in datasets, undersampling and oversampling. In his exploration, he also uses the MRN algorithm for the bracket problem of credit card fraud.

The problem of credit card fraud discovery has been anatomized with the Chebyshev Function Link Artificial Neural Network (CFANN). CFANN consists of two factors, functional expansion and literacy. Mukesh Kumar Mishra and Rajashree Gusto (10), authors who used CFANN to descry credit card fraud by comparing it with MLP, and the Decision Tree. MLP infers that the topology was structured into a number of layers. The first subcaste is called input subcaste, the middle subcaste is called the retired subcaste. This subcaste can have further than one subcaste, and the last subcaste is called the affair subcaste. Feed forward infers that all information flows in the same direction, the leftto-right direction, without intermittent links. Decision Tree is a structured tree that has a root knot and a number of internal and splint bumps. Their paper compares the performance of CFANN, MLP, and Decision Tree. The result of their study suggests that MLP outperforms CFANN and Decision Tree in fraud discovery. Again, CFANN makes accurate prognostications over the other two ways..

2.2 PROPOSED SYSTEM

The major aim of this project is to perform a comprehensive review of different fraud detection methods and some innovative machine learning techniques. Credit card is becoming more and more popular in financial transactions, at the same time frauds are also increasing. Conventional methods use rule-based expert systems to detect fraud behaviours, neglecting diverse situations, extreme imbalance of positive and negative samples. In this paper, a random forest-based fraud detection framework is proposed, to capture the intrinsic patterns of fraud behaviours learned from labelled data. Dataset of credit is fed to Machine learning algorithm for training. Transactions data is pre-processed to de-noise it while eliminating redundant data. Then important features are extracted and compare with dataset. Then using machine learning classifier credit card transactions are classified into fraud or authenticate one.

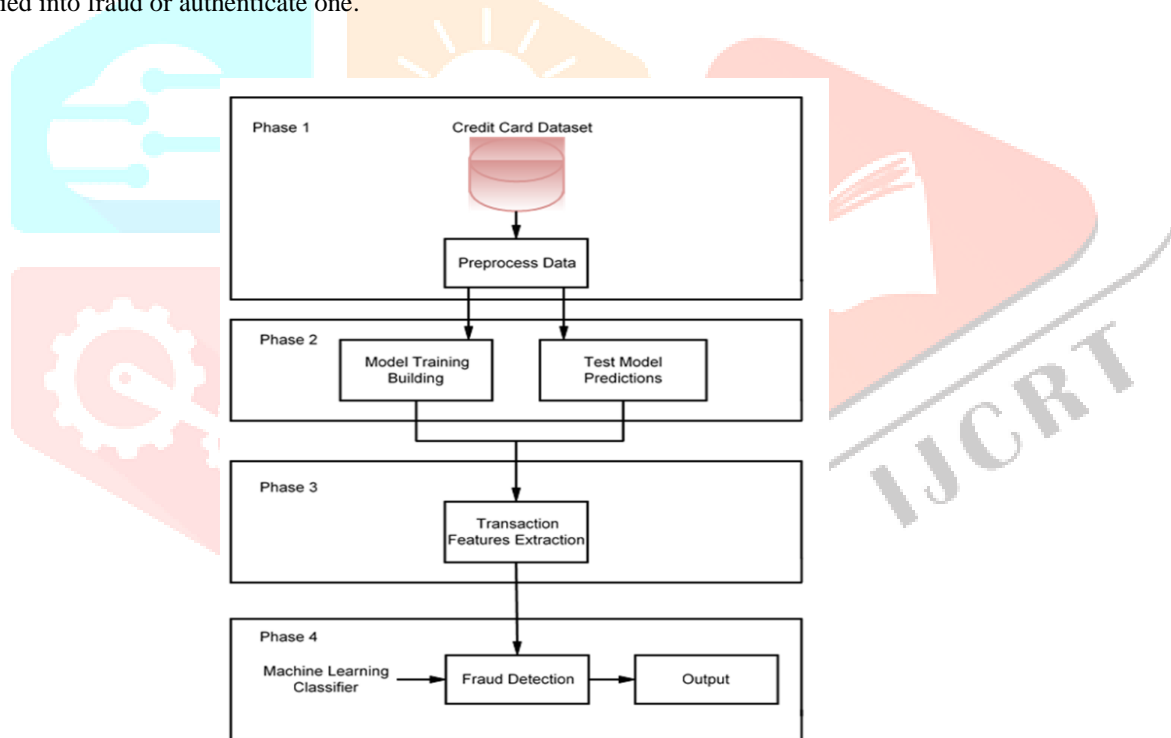


Fig 1: Block Diagram of Smart credit card Detection System

In the above diagram blocks used are: Dataset: Dataset of credit card transactions Training: Training a model simply means learning (determining) good values for all the weights and the bias from labeled examples. In supervised learning, a machine learning algorithm builds a model by examining many examples and attempting to find a model that minimizes loss; this process is called empirical risk minimization. Feature extraction: The extraction and matching of features is based on these measures. Besides the simple point feature a more advanced type of feature is also presented. Feature extraction technique is used to extract the features by keeping as much information as possible from large set of data of image.

Classification: Classification is a process of categorizing a given set of data into classes, It can be performed on both structured or unstructured data. The process starts with predicting the class of given data points. The classes are often referred to as target, label or categories. The classification predictive modeling is the task of approximating the mapping function from input variables to discrete output variables. The main goal is to identify which class/category the new data will fall into. Classification is performed using Machine learning algorithm. IV. CONCLUSION Credit card fraud detection is a fascinating domain. From this survey, we analyse that machine learning is the best compared to forecasting and classification. Machine learning techniques are mainly preferred in fraud detection, due to their high accuracy and detection rate. Even so, researchers find it difficult to achieve greater accuracy and detection speed. In addition, organizations are interested in finding ways to reduce costs and increase profits; you can find and select the method of previous studies. As a future work the work can be extended to online model. The use of online

learning will enable rapid detection of fraud cases, potentially in real-time. This in turn will help detect and prevent fraudulent transactions before they take place, which will reduce the number of losses incurred every day in the financial sector

2.3 CONCLUSION

Credit card fraud detection is a fascinating domain. From this survey, we analyse that machine learning is the best compared to forecasting and classification. Machine learning techniques are mainly preferred in fraud detection, due to their high accuracy and detection rate. Even so, researchers find it difficult to achieve greater accuracy and detection speed. In addition, organizations are interested in finding ways to reduce costs and increase profits; you can find and select the method of previous studies. As a future work the work can be extended to online model. The use of online learning will enable rapid detection of fraud cases, potentially in real-time. This in turn will help detect and prevent fraudulent transactions before they take place, which will reduce the number of losses incurred every day in the financial sector

REFERENCES

- [1] Samidha Khatri, Aishwarya Arora, Arun Prakash Agrawal, —Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison, 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Jan. 2020.
- [2] Hardik Manek, Sujai Jain, Nikhil Kataria, Chitra Bhole, Credit Card Fraud Detection Using Machine Learning International Journal of Innovative Research in Science, Engineering and Technology, Vol. 8, Issue 4, April 2019 IJIRSET DOI:10.15680/IJIRSET.2019.0804107 4507.
- [3] Jain R., Gour B., Dubey S., A hybrid approach for credit card fraud detection using rough set and decision tree technique, International Journal of Computer Applications 139(10) (2016).
- [4] Dermala N., Agrawal A.N., Credit card fraud detection using SVM and Reduction of false alarms, International Journal of Innovations in Engineering and Technology (IJJET) 7(2) (2016).
- [5] Tanmay Kumar Behera, Suvasini Panigrahi, Credit Card Fraud Detection: A Hybrid Approach Using Fuzzy Clustering and Neural Network, IEEE Computer Society, 2015
- [6] Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang, Credit Card Fraud Detection Using Convolutional Neural Networks, Springer International Publishing AG 2016.
- [7] Krishna Modi, Reshma Dayma Fraud Detection Technique in Credit Card Transactions using Convolutional Neural Network International Journal of Advance Research in Engineering, Science & Technology e-ISSN: 2393-9877, p-ISSN: 2394- 2444 Volume 4, Issue 8, August-2017
- [8] (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 1, 2018 18 | Page www.ijacsa.thesai.org Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine Apapan Pumsirirat, Liu Yan
- [9] N. Malini and Dr. M. Pushpa, —Analysis on credit card fraud identification techniques based on KNN and outlier detection in 3rd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB17)
- [10] Y. Lucas et al., —Towards automated feature engineering for credit card fraud detection using multiperspective HMMs, Future Generation Computer Systems, vol. 102, pp. 393–402, doi: 10.1016/j.future.2019.08.029, Jan. 2020.
- [11] E. Aji M. Mubarek, —Multilayer perceptron neural network technique for fraud detection, in International Conference on Computer Science and Engineering (UBMK), 2017.
- [12] S. P. Tanmay Kumar Behera, —Credit Card Fraud Detection: A Hybrid Approach using Fuzzy Clustering And Neural Network, in international conference on advances in computing and communication Engineering, 2015.
- [13] S. k. A. K. M. Ayushi agarwal, —Credit card fraud detection: A case study, in IEEE, New Delhi, India, 2015.0
- [14] S. Nami and M. Shajari, —Cost-sensitive payment card fraud detection based on dynamic random forest and k -nearest neighbors, Expert Systems with Applications, vol. 110, pp. 381–392, doi: 10.1016/j.eswa.2018.06.011, Nov. 2018.
- [15] S. Akila and U. Srinivasulu Reddy, —Cost-sensitive Risk Induced Bayesian Inference Bagging (RIBIB) for credit card fraud detection, Journal of Computational Science, vol. 27, pp. 247–254, doi: 10.1016/j.jocs.2018.06.009, Jul. 2018.
- [16] Maniraj, S & Saini, Aditya & Ahmed, Shadab & Sarkar, Swarna. Credit Card Fraud Detection using Machine Learning and Data Science. International Journal of Engineering Research and. 08. 10.17577/IJERTV8IS090031, 2019.
- [17] Vaishnavi Nath Dornadula, S Geetha, —Credit Card Fraud Detection using Machine Learning Algorithms, Procedia Computer Science, Vol. 165, pp.631–641, ISSN:1877-0509, https://doi.org/10.1016/j.procs. 2020.01.057, 2019.
- [18] A. Charleonnann, —Credit card fraud detection using RUS and MRN algorithm, in The 2016 Management and Innovation Technology International Conference (MITiCON-2016), 2016 © IEEE. doi: 978-1- 5090-4105-3/16/\$31.00
- [19] M. K. Mishra and R. Dash, —A comparative study of chebyshev functional link artificial neural network, multi-layer perceptron and decision tree for credit card fraud detection in 2014 13th International Conference on Information Technology, 2014 © IEEE. doi: 978-1-4799- 8084-0/14 \$31.00.
- [20] F. Carcillo, Y.-A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, and G. Bontempi, —Combining unsupervised and supervised learning in credit card fraud detection, Information Sciences, doi: 10.1016/j.ins.2019.05.042, May 2019.