iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** IV **Month of publication:** April 2019

DOI: https://doi.org/10.22214/ijraset.2019.4168

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

A Survey of Fraud Detection Techniques: Approach Oriented Perspective

Dr. Rahul Desai[1], Ajay Pratap Singh[2], Deepanshu Luhach[3], Nikhil Wagh[4], Prabhakar Singh[5]
[1, 2, 3, 4, 5]*Army Institute of Technology, Department of Information Technology* [1,2,3,4,5]*Savitribai Phule Pune University, Pune, Maharashtra*

Abstract: ***Financial fraud is a consistently developing threat with far consequences in the budgetary business. Data mining had assumed a basic job in the recognition of credit card extortion in online exchanges [1]. Credit card fraud detection is a data mining problem and it becomes challenging because of two noteworthy reasons - first, constant change in the profiles of normal and fraudulent behaviours and secondly, unbalanced credit card fraud data sets. Sampling approach on dataset, selection of variables and detection technique(s) used greatly affects the performance of fraud detection in credit card transactions [2]. Many techniques have been proposed to confront the issue. However, all of these techniques have the same goal of attaining better results, each one has its own drawbacks, advantages and characteristics [3]. This paper explores the challenges of credit card fraud and surveys different fraud detection techniques, datasets and assessment criteria.The points of interest and inconveniences of various strategies are specified and looked at. Consequently, open issues for credit card fraud detection are gathered and examined as rules for new analysts.***

Index Terms: Naive Bayes, Logistic Regression, Gaussian Distribution, Neural Network

I. INTRODUCTION

At the current state of the world, financial organizations expand the availability of financial facilities by employing of innovative services such as credit cards, Automated Teller Machines (ATM), internet and mobile banking services. In addition, alongside the quick advances of web based business, the utilization of charge card has turned into a comfort and vital piece of money related life. Credit card is a payment card supplied to customers as a system of payment.
There are lots of advantages in using credit cards such as:

A. Ease of purchase
B. Keep customer credit history
C. Protection of Purchases

Disregarding all advantages, the problem of fraud is a significant issue in e-banking administrations that compromise credit card transactions particularly. Fraud is a violation of public law in which the fraudster gains an unethical advantage or causes unlawful damage. The estimation of amount of damage made by fraud activities indicates that fraud costs a very considerable sum of money. Credit card fraud is increasing significantly with the development of modern technology [3].
Credit card fraud is a massive problem for e-commerce retailers. The original cardholder will call their bank to reverse the transaction(s) when a stolen credit card is used, that initiates a chargeback. Merchants bring about the risk for chargebacks, restoring the cash, miss out on sent products, and pay additional expenses to their payment processors. Having a chargeback rate over 1% causes extra fees, assessments, and eventually termination by the card networks. Fraud has many origins - black markets of stolen cards, card numbers obtained during a major hack, sophisticated credit card rings, petty thieves, and more [4]. Although, fraud detection has gained attention and extensively studied especially in recent years, but there are not many surveys about the issues in fraud detection, neither analyze the advantages and disadvan-tages of various techniques based on approach oriented perspective. Therefore in this paper, we attempt to collect and integrate a complete set of evaluation employed criterions in literature and analyze them from various aspects.
The main contributions of this work are highlighted as follows :

1) To the best of our knowledge, the absence of complete and detailed survey is an important issue, which is addressed by analyzing the state of the art in credit card fraud detection
2) The state of the art techniques are described and classified from different aspects

©IJRASET: All Rights are Reserved

3) There is no standard or benchmark to evaluate detection methods. We attempt to gather different data investigated by researchers, extract common attributes and evaluate them on common basis

The rest of the paper is organized as follows: existing platforms and general description of different techniques is presented, challenges of fraud detection are identified, advantages and disadvantages of various techniques are classified, dataset insights, approach and evaluation criteria is determined. At last open issues of credit card fraud detection are exhibited.

II. EXISTING PLATFORMS

A. Overview

Financial fraud has it's far reaching consequences in the finance industry, corporate organizations, and gov-ernment. Fraud can be characterized as criminal trickiness with expectation of getting monetary benefit. High dependence on internet technology has enjoyed increased credit card transactions. As credit card transactions become the most prevailing mode of payment for both online and offline transaction, credit card fraud rate also accelerates.

B. Classification of Fraud

Credit card companies needs to perceive false exchanges with the goal that clients are not charged for things that they didn't buy. We can distinguish frauds among two families:

1) *Inner card fraud or Application fraud:* Occurs as a result of consent between cardholders and bank by using false identity to commit fraud

2) *External card fraud or Behavioral fraud:* Involves the use of stolen credit card to get cash through dubious means

III. GENERAL DESCRIPTION OF DIFFERENT TECHNIQUES

A. Naive Bayes

Naive Bayes is a classification algorithm. It is used for binary (two-class) and multi-class classification problems and is easiest to understand when described using binary or categorical input values. It is called Naive Bayes because the calculation of the probabilities for each hypothesis are simplified to make their calculation tractable. Rather than attempting to calculate the values of each attribute value, they are assumed to be conditionally independent given the target value. This is an exceptionally solid presumption that is most improbable in genuine information, i.e. that the attributes do not interact. Nevertheless, the methodology performs surprisingly well on information where this presumption does not hold. It's primary disadvantage is that it can't learn interactions between features [2][5][6].

B. Logistic Regression

To conduct analysis when the dependent variable is binary, logistic regression is a fitting approach. Logistic regression is a predictive analysis. It is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. It provide lots of ways to regularize your model, and unlike Naive Bayes you don't have to worry as much about your features being correlated. Use it in the event that you need a probabilistic structure (e.g., to effectively alter grouping limits, to state when you're uncertain, or to get certainty interims) or in the event that you hope to receive more training data in the future for which you need to have the capacity to rapidly consolidate into your model [2].

C. Gaussian Distribution

Gaussian distribution (Normal distribution) is very common by nature. Almost all variables are distributed approximately normally. Incredible number of procedures in nature and sociologies normally pursues the Gaussian distribution. Even when they don't, the Gaussian gives the best model estimation for these procedures. In spite of the fact that they are just roughly ordinary, they are commonly very close. The idea behind it is central limit theorem. Central limit theorem states that when we include extensive number of autonomous irregular factors, independent of the first circulation of these factors, their standardized aggregate tends towards a Gaussian distribution. Numerous sort of factual tests are gotten from normal distribution and furthermore function admirably if the distribution is roughly ordinary. A few tests functions admirably even with wide de-viation from normality. It is characterized by just two numbers: mean μ and variance σ^2 or standard deviation σ [7].

©IJRASET: All Rights are Reserved

D. Neural Network

Neural network is a set of interconnected nodes designed to imitate the functioning of the human brain. Every node has a weighted association with a few different nodes in neighboring layers. Individual node takes the input from associated nodes and utilize the weights together with a basic capacity to compute output values. Neural systems come in numerous shapes and models. The Neural network architecture, including the number of hidden layers, the number of nodes within a specific hidden layer and their connectivity, must be specified by user based on the complexity of the problem. It has the ability to learn and model non-linear and complex relationships, which is really important because in real life, many of the relationships between inputs and outputs are non-linear as well as complex. In the wake of gaining from the underlying information sources and their connections, it can surmise concealed connections on inconspicuous information also, along these lines influencing the model to sum up and anticipate on concealed information [8][9][10][11].

IV. DATASET, APPROACH AND EVALUATION

A. Dataset

The lack of publicly available database has been a limiting factor for the publications on financial fraud detection, particularly credit card transactions. On the other hand, credit card is inherently private so, creating a proper data set for this purpose is very difficult and there are no standard techniques to do this. Also, there is no universal corpus for credit card [2][3].

A legitimate data set is a data set which covers different misrepresentation and a few properties of customer profile or conduct. We believe that the contribution of attributes is a critical factor that should be considered. Also, a proper data set should be able to reflect the real world of credit card [12]. Primary attributes are attributes of credit card transactions which are available in the most datasets. Some of the common attributes in most datasets are as follows:

1) Posting Date - date when transaction was posted to the accounts
2) Merchant Category Code (MCC) - code devoted to each good
3) Transaction Date and Time - date and time at which the transaction was actually performed
4) Transaction Status - status of transaction success/fail
5) Transaction Place - place of transaction (usually determined by IP Address)
6) Money Amount - amount of Money
7) Transaction Type - type of transaction payment/deposit/transfer etc
8) Customer Identification - identification code which advocate to each customer
9) Scheme - type of credit card used, e.g. MasterCard, Visa etc

B. Approach

In general, credit card fraud data is highly imbalanced. There are a few different ways to approach this clas-sification problem contemplating this unbalance.

1) Collect more data
2) Changing the performance metric
a) Use the confusion matrix to calculate Precision, Recall
b) F1score (weighted average of precision recall)
c) Use Kappa - which is a classification accuracy normalized by the imbalance of the classes in the data ROC curves - calculates sensitivity/specificity ratio
3) Resampling the dataset
a) This is a method that will process the data to have an approximate 50-50 ratio
b) One way to achieve this is by OVER-sampling, which is adding copies of the underrepresented class (better when you have little data) - positive class (fraud) is over-sampled
c) Another is UNDER-sampling, which deletes instances from the over-represented class (better when one has lots of data) - negative class (legitimate) is under-sampled

C. Evaluation

There are an assortment of measures for different algorithms and these measures have been created to assess altogether different things. False Positive (FP), False Negative (FN), True Positive (TP), and True Negative (TN) and the relation between them are quantities which usually adopted by credit card fraud detection re-searcher to compare the accuracy of different approaches. The aim of all algorithms and techniques is to minimize FP and FN rate and maximize TP and TN rate and with a good detection rate at the same time.

©IJRASET: All Rights are Reserved

V. OPEN ISSUES

A lot of researches have been devoted to detection of external card fraud which accounts for majority of credit card frauds. Detecting fraudulent transactions using traditional methods of manual detection is time consuming and inefficient, thus the advent of big data has made manual methods more impractical. Machine Learning techniques are one of the notable methods used in solving credit fraud detection problem.

While fraud detection has gained wide-scale attention in the literature, there are yet some issues (a number of significant open issues) that researchers face and have not been addressed before adequately. We hope this overview focuses the direction of future research to provide more efficient techniques for credit card fraud detection.

These issues are as follows

A. Nonexistence of standard and comprehensive credit card benchmark or dataset
B. Nonexistence of standard algorithms
C. Nonexistence of suitable metrics
D. Lack of adaptive credit card fraud detection systems

VI. CONCLUSION

To improve merchant's risk management level in a programmed, logical and compelling way, constructing a precise, accessible and easy handling credit card risk monitoring system is one of the key tasks for the merchant banks. Performance of classifiers varies across different evaluation metrics. This study suggests hybrid sampling approach on binary classification of imbalanced data and evaluation metrics to be used.

This survey paper is all about the different techniques and approaches for credit card fraud detection. There have been significant research done to overcome the problems of the techniques used earlier and a better approach has been proposed.

Expected future areas of research could be in examining meta-classifiers and meta-learning approaches in handling highly imbalanced credit card fraud data.

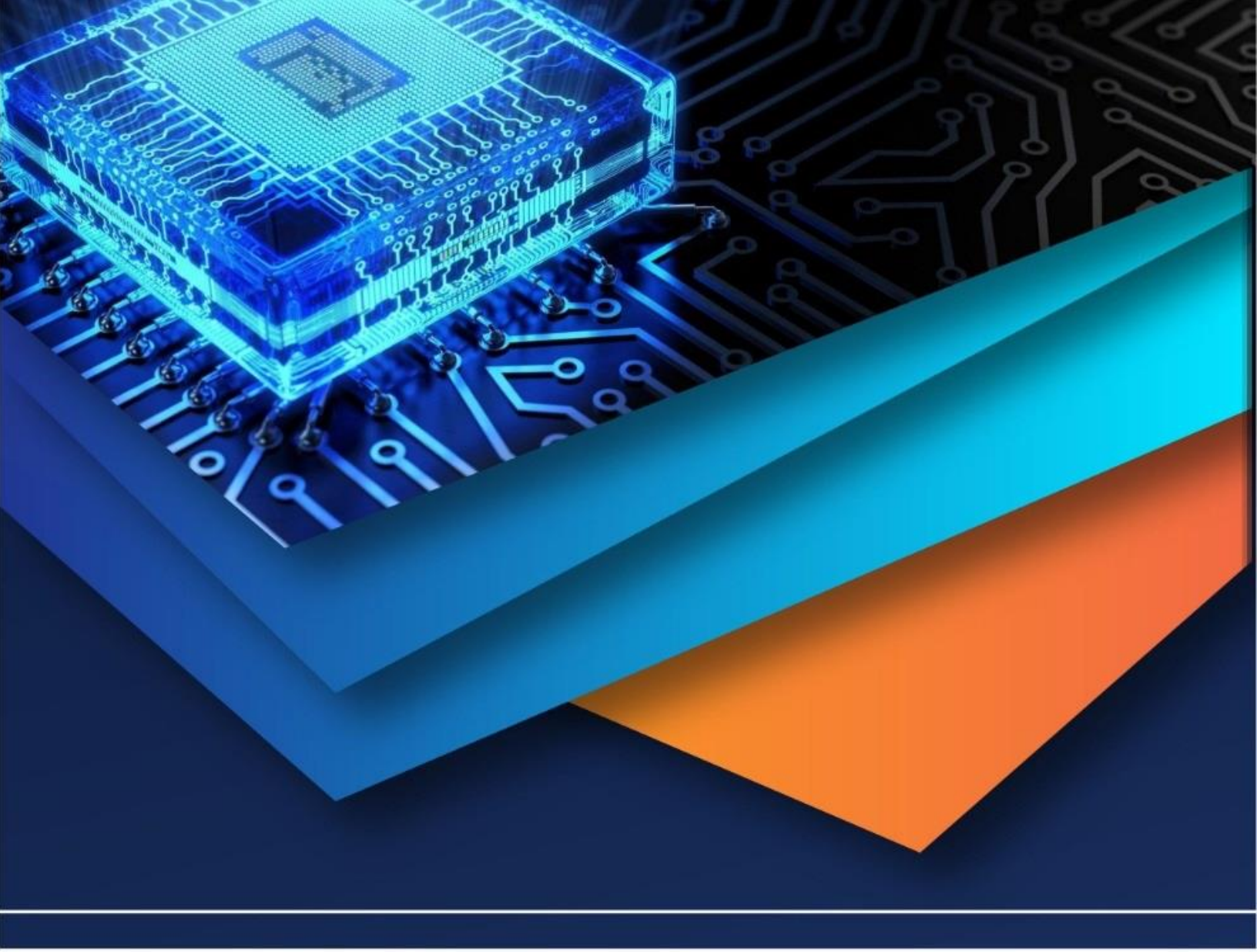
VII. ACKNOWLEDGEMENT

We would like to express our deepest appreciation to all those who provided us the possibility to complete this survey. We would like to thank our college Army Institute of Technology for supporting us and providing us with various resources. We give special gratitude to our HOD, Dr. Sangeeta Jadhav, whose contribution in stimulating suggestions and encouragement, helped us to coordinate our survey.

REFERENCES

[1] Jurgovsky, Johannes Granitzer, Michael Ziegler, Konstantin Calabretto, Sylvie Portier, Pierre-Edouard He, Liyun Caelen, Olivier, "Sequence Classification for Credit-Card Fraud Detection". In:Expert Systems with Applications, 2018.
[2] John O Awoyemi, Adebayo O Adetunmbi, Samuel A Oluwadare, "Credit card fraud detection using Machine Learning Techniques : A Comparative Analysis". In:Institute of Electrical and Electronics Engineers, 2017.
[3] Samaneh Sorournejad, Zahra Zojaji, Reza Ebrahimi Atani, Amir Hassan Monadjemi, "A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective". In:Institute of Electrical and Electronics Engineers, 2015.
[4] AlkaHerenj, Susmita Mishra, "Secure Mechanism for Credit Card TransactionFraud Detection System". In:International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 2, February 2013.
[5] Luis Jose S Santos, Shirlee R Ocampo, "Bayesian Method with Clustering Algorithm for Credit Card Transaction Fraud Detection". In:Romanian Statistical Review, 2018.
[6] Milgo, Carolyne, "A Bayesian Classification Model for Fraud Detection over ATM Platforms". In: IOSR Journal of Computer Engineering, Volume 18, Issue 4, Jul-Aug 2016.
[7] E Duman, A Buyukkaya, I Elikucuk, "A novel and successful credit card fraud detection system implemented in aturkish bank". In:Data Mining Workshops, IEEE 13th International Conference, 2013, pp. 162-171.
[8] A Roy, J. Sun, R Mahoney, L Alonzi, S Adams, P Beling, "Deep learning detecting fraud in credit card transactions". In:Systems and Information Engineering Design Symposium, 2018, pp. 129-134.
[9] Chandrahas Mishra, Dharmendra Lal Gupta, Raghuraj Singh, "Credit Card Fraud Identification Using Artificial Neural Networks". In:International Journal of Computer Systems, Volume 04, Issue 07, July 2017.
[10] Aman Gulati, Prakash Dubey, Md Fuzail C, Jasmine Norman, Mangayarkarasi R, "Credit card fraud detection using neural network and geolocation". In:14th ICSET-2017 IOP Conf. Series: Materials Science and Engineering, 2017.
[11] Tao Guo, Gui-Yang Li, "Neural data mining for credit card fraud detection". In:Institute of Electrical and Electronics Engineers, 2008.
[12] Bahnsen, Alejandro Correa, Djamila Aouada, Aleksandar Stojanovic and Bjorn E Ottersten, "Feature engineering strategies for credit card fraud detection". In:Expert Syst. Appl. 51 (2016): 134-142.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)