A Survey on Credit Card Fraud Detection and Prevention Using Hybrid Algorithm

Tripti Goutam [1], Ghanshyam Sahu [2], Lalit Kumar P Bhaiya [3]

[1]M. Tech Scholar, Department of CSE, Bharti College of Engineering and Technology, Durg, Chhattisgarh, India.

[2]Department of Computer Science & Engineering, Bharti College of Engineering and Technology, Durg, Chhattisgarh, India.

[3]Bharti University, Durg, Chhattisgarh, India.

Corresponding Author: just4tripti@gmail.com

Abstract: - Fraud related to credit cards is increasing as it becomes the most popular method of payment for both ordinary purchases and those made online. Financial institutions and service providers are facing a significant financial burden as a result of the rising number of electronic payments, which is requiring them to continuously enhance their fraud detection systems. Nonetheless, although being widely used in other fields, contemporary data-driven and learning-based methods are still making moderate progress in business applications. In this paper we are representing the types of credit frauds and its detections by reviewing various published papers and its different methods of solving by various algorithms. This essay contrasts and evaluates a few effective methods for identifying credit card fraud. The approaches used to detect credit card fraud are the Dempster Shafer and Bayesian Learning Fusion, Hidden Markov Model, Artificial Neural Networks and Bayesian Learning Approach BLAST and SSAHA Hybridization, and Fuzzy Darwinian System.[1] A description of these methods is provided in Section II. Part III provides a comparison of such strategies, and Section IV provides a summary of fraud detection methods.

Key words: Credit Card, Fraud Detection, Fraud Detection Framework, Supervised and Unsupervised Techniques.

I. INTRODUCTION

Credit Card- It is an instrument for the transactions, even if cardholders don't have the necessary funds in their bank account, they can still make transactions using a credit card, which is a cashless payment method. A charge card, also demands of balance to be paid in full every month or as a conclusion of each statement cycle, that is distinct from a typical credit card. Credit cards, on the other hand, give users the option to accrue ongoing debt that is subject to interest charges.[2]

Manuscript revised April 19, 2023; accepted April 20, 2023. Date of publication April 23, 2023.
This paper available online at www.ijprse.com
ISSN (Online): 2582-7898; SJIF: 5.59

The fact that a credit card often involves a third member or third party which helps to pays the seller and will be reimbursed by the buyer, it opposed a charge card, only for postpones payment of buyers until a later time, and another way is that credit cards and charge cards were differ from one another.

Credit Card Fraud - A significant surge in fraudulent activity has been caused by a fast increase in credit card transactions. For theft and fraud committed using a credit card as a fraudulent source of money in a specific transaction, credit card fraud is a broad word. Many different methods are used by credit card fraudsters to commit fraud.[3] It's crucial to first comprehend the mechanisms of credit card fraud detection in order to combat it effectively. Owing to several mechanisms for detecting and preventing credit card fraud, it has significantly decreased over time.

Fraud can be defined as criminal deception carried out with the purpose of making money or say as profit. Increasing reliance on internet technologies has resulted in an increase in credit card transactions. As credit card transactions replace cash as the predominant mode of payment for both online and offline, so the rate of fraud increasing rapidly with both types. There are two types of credit card fraud: internal and external. While external card fraud entails using a stolen credit card to obtain money through shady ways, inner card fraud happens as a result of agreement between cardholders and the bank and involves using a fictitious identity to commit fraud. The bulk of credit card frauds are external card fraud, which has been the subject of much investigation.

In the future we implement the fraud detection system by using mock dataset from dataset sampling strategy, variable choice, and detection method(s) employed all have a significant impact on the effectiveness of fraud detection in credit card transactions. The effectiveness of naive bayes, k-nearest neighbor, and logistic regression on highly skewed credit card fraud data is examined in this research. Dataset of credit card transactions with 284,807 transactions is obtained from European cardholders. On the skewed data, a hybrid technique of under-sampling and over- sampling is used. So that we can able to develop credit fraud system which will be helpful for any normal person which is using this system.

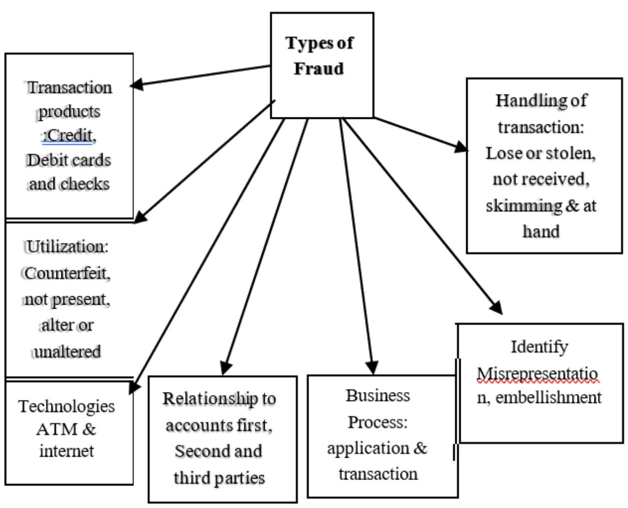

Fig.1. Types of Fraud

II. Related work

Credit card transaction classification is primarily a binary classification problem. Here, a credit card transaction falls into one of two categories: either it was fraudulent or it was legal (positive class). The target of fraud detection is clearly and typically seen in the data mining classification challenge, where the classification of credit card transactions as legal or fraudulent must be done correctly.[4]

2.1 Credit Card Fraud

Internal and external fraud involving credit cards have been divided into two categories, while fraud involving traditional cards (application, stolen, account takeover, fake, and counterfeit), fraud involving merchants (merchant collusion and triangulation), and fraud involving the Internet (site cloning, credit card generators, and false merchant sites) have been divided into three categories. According to [5], the overall amount of fraud losses suffered by banks and businesses worldwide in 2014 was over USD 16 billion, an increase of around USD 2.5 billion above losses incurred the year before. This means that every USD 100 contains 5.6 cents of fraudulent activity.

2.2 Credit Card Fraud Detection

The rate of fraud tends to increase as credit cards become the most widely used form of payment (for both online and offline purchases). Using conventional methods to identify fraudulent transactions.

The development of big data has made manual detection approaches more impractical because they are time-consuming and imprecise. Financial institutions have, nevertheless, resorted to clever methods. These sophisticated fraud strategies are based on computational intelligence (CI). Methods for statistically detecting fraud have been split into two categories: both closely and loosely watched [6]

In way to categorizes any new transactions as fraudulent or legitimate in a supervised fraud detection method, there are models that will be estimated based on prototype of fraudulent and legitimate transactions, where in unsupervised fraud detection, outliers' transactions are identified as potential examples of fraudulent transactions. [7] has a thorough overview of supervised and unsupervised approaches. Studies on a variety of strategies have been conducted in an effort to solve the problem of credit card fraud detection. Neural network models (NN), Bayesian networks (BN), intelligent decision engines (IDE), expert systems, meta-learning agents, machine learning, pattern recognition, rule-based systems, logic regression (LR), support vector machines (SVM), decision trees, k-nearest neighbor (KNN), meta learning strategy, adaptive learning, etc. are a few detection techniques which are presented.

2.3 Feature (Variables) selection

The examination of a cardholder's spending patterns is the cornerstone of credit card fraud detection. The ideal combination of characteristics that best reflect the distinctive behavior of a credit card are used to analyses this spending profile. Both a valid and fraudulent transaction's profile tends to change over time. To effectively classify credit card transactions, it is necessary to choose factors that significantly distinguish both profiles. The factors that make up the card usage profile and the methods utilized have an impact on how well credit card fraud detection systems work. These characteristics are derived from a credit card's transaction history and previous transaction history. These variables fall under the categories of total transactions statistics, geographical statistics, merchant type statistics, and time-based amount statistics, which make up the five primary variable kinds.

2.4 Fraud Detection Framework

It is divided into training and prediction components. Four sections make up the bulk of the training portion: feature engineering, sample techniques, feature transformation, and a based training procedure. While the prediction portion is online, the training portion is offline. When a transaction occurs, the prediction component can quickly determine if it is legitimate or fraudulent. The feature extraction, feature transformation, and classification modules make up the detection process.

In order to represent more complex consumption patterns, we apply aggregation methodologies for feature extraction and add trade entropy to the set of classical features. In the typical data mining procedure, feature engineering is followed by model training. Yet, there is a problem with the credit card data being so unbalanced. For the purpose of creating fake frauds, we provide a cost-based sampling technique.[8]

2.5 Feature Engineering for Temporal Sequence

For accurate categorization while building a credit card fraud detection system, features must be carefully chosen. It should come as no surprise that a lot of research has gone into creating expressive features. The set of attributes, which includes elements like time, amount, merchant category, account number, transaction type, etc., is relatively consistent across numerous credit-card datasets. These characteristics are utilised directly as input features to train a classifier on the binary classification issue in a "conventional" fraud detection system. Such systems function at the transaction level, treating transactions as isolated events, and therefore fail to take into account the fact that the frequency or volume of transactions over time and among merchants can provide a wealth of information about a certain account.

III. Literature Review

Padvekar SA, Kangane PM, Jadhav KV [9] – They revealed that hidden-markov models are routinely used to identify credit card misrepresentation during transactions. High misrepresentation inclusion is combined with a low false alarm rate for the concealed mark-ovmodel. They utilised the scopes of exchange amount as the perception images, while the classifications of products are contemplated to be conditions of the HMM. They devised a method for determining cardholders' spending patterns, and used this information to estimate the model parameters and determine the value of the observation symbols. It is also described how HMM can determine whether an upcoming transaction is fraudulent or not. According to relative analyses, the accuracy of the framework is on the thin edge of 80% over a wide range of data.

Xuan S, Liu G, Li Z, Zheng L, Wang S, Jiang C [10], The preparation set for the two algorithms—random-tree-based random forests and CART-based random forests—comes from bootstrapped experiments. Using various datasets with various dataset proportions, the three experiments were run for the two methods. For each of the three experiments, the effectiveness of these algorithms was evaluated, and the metrics of intervention rate of transaction and average rate of model were included. In each test, the cart-based random forest outperformed the baseline model.

Fernandes, E. R., & de Carvalho, A. C [11] they address the disadvantage of discarding majority samples in the overlap zone and suggests an evolutionary ensemble- based approach using classifier ensembles to address this issue. The fundamental classifier is still trained using a preprocessed dataset, where the vast majority of the samples are removed because they overlap. And the overlapping region is found using the neighborhood-based minimal spanning tree. Despite the fact that these neighborhood- based approaches may immediately and precisely identify the overlapping subset, their high calculation costs make them inapplicable when the actual data is high-dimensional and contains a large number of samples.

Das, S., Datta, S., & Chaudhuri, B. B [12] it gives a thorough investigation of data abnormalities, including class imbalance, tiny disjuncts, class skew, and missing features, and their relationships. They also list the notable and most recent approaches to dealing with data irregularities. In order to encourage better model design, a few noteworthy future research directions are suggested. Together with an investigation of the connections between overlap and class imbalance, possible solutions to the challenging issue of class imbalance with overlap are also put forth.

Lee, H. K., & Kim, S. B [13] proposes a solution to this issue based on overlap sensitive margin (OSM). An original dataset is divided into soft- and hard-overlap subgroups using a modified fuzzy support vector machine. Next, the support vector machine technique and the one- nearest neighbor approach are used, respectively, to classify the two obtained subsets. These techniques employ an under-sampling technique on the overlapping subset to achieve a distinct boundary biassing in favor of the minority class, however it is difficult to determine which majority sample should be deleted to prevent uncontrollably losing information.

Zhou, C., & Paffenroth, R. C. [14], in this paper it is to understand the basic characteristics of the minority samples (fraud transactions). The majority samples and some of the minority samples (legitimate transactions) that heavily overlapped with these minority samples are combined with the aid of this profile to form the overlapping subset, while the remaining majority samples and minority samples (fraudulent transaction anomalies) form the nonoverlapping subset. Contrary to the typical notion that anomaly detection models use a majority sample profile and treat anomalies as minority samples, this unique proposal uses anomaly detection models to identify anomalies based on the profile of the minority class.

IV. Conclusion

In this paper, we discussed and review different research papers which makes credit card fraud is a dishonest criminal conduct. Recent developments in the credit card industry were covered in this article. This essay described the various fraud kinds and offered ways to spot them, including bankruptcy fraud, counterfeit fraud, theft fraud, application fraud, and behavioural fraud. For detection some measures will be taken by the help of algorithms like genetic algorithm, neural networks, clustering techniques etc which helps to resolve our problem. Building scoring models that predict fraudulent conduct while taking into account the behavioural domains related to the various types of credit card fraud highlighted in this research will be the key tasks, along with assessing the accompanying ethical ramifications.

References

[1]. Abhinav Srivastava, Amlan Kundu, Shamik Sural, Arun K. Majumdar, “Credit Card Fraud Detection using Hidden Markov Model,” IEEE Transactions on Dependable and Secure Computing, vol. 5, Issue no. 1, pp.37-48, January-March 2008.

[2]. Schneider, Gary (2010). Electronic Commerce. Cambridge: Course Technology. p. 497. ISBN 978-0-538-46924-1.

[3]. S. Benson Edwin Raj, A. Annie Portia, “Analysis on Credit Card Fraud Detection Methods”, International Conference on Computer, Communication and Electrical Technology – ICCCET2011, 18th & 19th March, 2011.

[4]. Seeja, K. R., and Zareapoor, M., (2014). FraudMiner: A Novel Credit Card Fraud Detection Model Based on Frequent Itemset Mining, The Scientific World Journal, Hindawi Publishing Corporation, Volume 2014, Article ID 252797, pp. 1 – 10.

[5]. The Nilson Report. (2015). U.S. Credit & Debit Cards 2015. David Robertson.

[6]. Bolton, R. J. and Hand, D. J., (2001). Unsupervised profiling methods for fraud detection, Conference on Credit Scoring and Credit Control, Edinburgh.

[7]. Kou, Y., Lu, C-T., Sinvongwattana, S. and Huang, Y-P., (2004). Survey of Fraud Detection Techniques, In Proceedings of the 2004 IEEE International Conference on Networking, Sensing & Control, Taipei, Taiwan, March 21-23.

[8]. Bhattacharyya, S., Jha, S., Tharakunnel, K., Westland, J.C.: Data mining for credit card fraud: a comparative study. Decis. Support Syst. 50(3), 602–613 (2011).

[9]. Padvekar SA, Kangane PM, Jadhav KV (2016) Credit card fraud detection system. Int J Eng Comput Sci 5(4):16183–16186

[10]. Xuan S, Liu G, Li Z, Zheng L, Wang S, Jiang C (2018) Random Forest for credit card fraud detection. In: ICNSC 2018—15th IEEE International conference on networking, sensing and control, pp 1–6.

[11]. Fernandes, E. R., & de Carvalho, A. C. “Evolutionary inversion of class distribution in overlapping areas for multi-class imbalanced learning”. Information Sciences, 494, 141–154.

[12]. Das, S., Datta, S., & Chaudhuri, B. B., “Handling data irregularities in classification: Foundations, trends, and future challenges. Pattern Recognition”, 2018, 81, 674–693.

[13]. Lee, H. K., & Kim, S. B. “An overlap-sensitive margin classifier for imbalanced and overlapping data. Expert Systems with Applications”, 2018, 98, 72–83.

[14]. Zhou, C., & Paffenroth, R. C. “Anomaly detection with robust deep autoencoders”, In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017 (pp. 665–674).