IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, April 2024

Credit Card Fraud Detection

Prof. Dipali Dube[1], Siddhesh Kharge[2], Abhay Nighot[3], Omkar Fulsundar[4], Prasad Naykodi[5]
Guide, Department of Computer Engineering[1]
Students, Department of Computer Engineering[2,3,4,5]
Vidya Niketan College of Engineering, Ahmednagar, Maharashtra, India

Abstract: *The issue of credit card fraud presents a notable concern within the financial sector, leading to considerable financial losses for both financial institutions and consumers. To address this challenge, this study investigates the application of machine learning methods for detecting credit card fraud. We explore the performance of diverse ma- chine learning algorithms on an actual dataset and suggest an ensemble-based approach that harnesses the strengths of multiple models. Our experimental outcomes demon- strate the effectiveness of machine learning in accurately identifying fraudulent transactions while minimizing false positives.*

Keywords: credit card fraud.

I. INTRODUCTION

In the digital era, the widespread adoption of credit cards for financial transactions has transformed the way we han- dle payments and conduct business. While this conve- nience has undoubtedly enhanced our daily lives, it has also brought forth a significant challenge: credit card fraud. Activities such as unauthorized transactions, identity theft, and card-not-present fraud have surged, presenting sub- stantial financial risks to financial institutions and consumers alike. Detecting and preventing credit card fraud is cru- cial in today's financial landscape, necessitating a proac- tive and adaptable approach capable of swiftly identifying fraudulent transactions while minimizing disruptions to le- gitimate cardholders. Traditional rule-based systems have limitations, often struggling to keep pace with the evolving tactics of fraudsters. In contrast, machine learning offers a promising solution, harnessing the power of data analytics and predictive modeling to differentiate between genuine and fraudulent transactions.

This paper provides an extensive review of prior work, tracing the evolution of credit card fraud detection meth- ods and underscoring the pivotal role of machine learning in advancing this field. We delve into the intricacies of data preprocessing, highlighting the significance of ade- quately preparing datasets to extract meaningful insights. Additionally, we outline our methodology, encompassing the selection of machine learning algorithms, feature en- gineering, and the development of an ensemble-based ap- proach that amalgamates the strengths of multiple models. Our experiments and results showcase the practical effi- cacy of these techniques. We assess our models using var- ious performance metrics, including accuracy, precision, recall, F1-score, and ROC curves, offering a comprehen- sive evaluation of their capabilities. Our findings not only demonstrate the accuracy of our machine learning models but also emphasize the importance of striking a balance between fraud detection and minimizing false positives, a critical consideration in the financial sector.

II. LITERATURE REVIEW

Previous research has explored a plethora of methods to tackle fraud detection, ranging from supervised and unsu- pervised approaches to hybrid ones. Consequently, under- standing the technologies associated with credit card fraud detection and gaining insights into various fraud types have become imperative. Over time, the evolution of fraud pat- terns has introduced novel forms of fraudulent activities, thereby piquing the interest of researchers. The subse- quent part of this section elaborates on individual machine learning algorithms, models, and fraud detection systems utilized in fraud detection efforts. Issues identified during the review process have been analyzed for potential incor- poration into future implementations of efficient machine learning models.

Past studies have highlighted several issues in fraud detection. Some papers, like and note the scarcity of real- life data due to privacy concerns, making it challenging to obtain authentic datasets. Others, such as and discuss the problem of imbalanced data, where there are significantly fewer fraud cases compared to non-fraudulent transactions. Dealing with

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, April 2024

large datasets can also be time-consuming, as mentioned in, due to the computational demands of data mining techniques. Additionally, overlapping data poses challenges, as legitimate transactions may resemble fraud- ulent ones, and vice versa, as pointed out . Handling cat- egorical data is another hurdle, as most machine learning algorithms do not support this type of data. Finally, choos- ing appropriate detection algorithms and selecting relevant features are cited as challenges in detecting frauds, as train- ing these algorithms can be time-intensive compared to prediction tasks.

III. EXPERIMENTAL METHODOLOGY

A. Data Overview

This dataset was compiled by merging two primary sources: the fraud transactions log file and the comprehensive trans- actions log file. The fraud transactions log file contains records of all instances of online credit card fraud, whereas the all transactions log file encompasses all transactions logged by the respective bank over a specified timeframe.

To preserve confidentiality, certain sensitive attributes such as card numbers were encrypted using hashing techniques, as per the agreement between the bank and the authors of the study.Upon examining the combined dataset, it became evident that the distribution of data was heavily skewed due to the disproportionate numbers of legitimate transactions and fraudulent occurrences. Specifically, the file contain- ing fraud cases comprised 200 records, whereas the trans- action log file contained 917,781 records.

B. Data Preparation

Initially, the raw data were segmented into four distinct datasets based on identified fraud patterns. This segmenta- tion was informed by insights provided by the bank. The four datasets are as follows:

Transactions associated with Risky Merchant Category Codes (MCC). Transactions exceeding 10000rs in value. Transactions flagged with risky ISO Response codes. Trans- actions originating from unknown web addresses. These four datasets were then utilized in two distinct approaches:

Type A: Transforming the raw data into numerical rep- resentations. Type B: Categorizing the raw data without any transformation. For datasets 1, 2, and 3, Type A data preparation methodology was applied, while Type B was applied solely to dataset number 4. The data preparation process typically involves cleaning, transformation, inte- gration, and reduction of data. In the case of Type A prepa- ration, all of these steps were applied to the first three datasets to prepare them numerically. However, for cate- gorical data preparation (Type B), all steps except for data transformation were implemented. The fundamental steps involved in Type A data preparation are outlined below:

Data Cleaning:

Filling in missing values is a crucial aspect of the data cleaning process. Various approaches exist to address this issue, such as excluding entire tuples, but many of these methods are prone to introducing biases into the data. How- ever, since the source file containing genuine transactions did not have any records with missing values, this was not a concern. Tuples containing meaningless values were re- moved from the files as they do not contribute to generat- ing meaningful data and could potentially bias the dataset. Furthermore, additional changes were made, including re- moving unnecessary columns and splitting the datetime column into two separate components.

Data Integration:

Prior to implementing further modifications, the two data sources were integrated because the fraudulent and genuine record files were stored separately. Figure 1 illus- trates the mapping process employed to merge the datasets together.

Data Transformation:

In this stage, all categorical data were unified into a comprehensible numerical format. The transactional dataset encompasses various data types with diverse ranges. Con- sequently, data transformation involved normalization, which scales the attribute data to fit within a smaller numeric range.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, April 2024

Data Reduction:

The chosen strategy for data reduction is dimensional- ity reduction. This approach aims to mitigate the risk of learning incorrect data patterns, ensuring that the selected features effectively eliminate irrelevant aspects and charac- teristics of the fraud domain, as highlighted in . Principal Component Analysis (PCA) is a widely recognized method for dimensionality reduction. By applying PCA, the fea- ture selection issue is addressed from a numerical analysis standpoint. PCA effectively selects features by identifying the appropriate number of principal components.

IV. CONCLUSION

Credit card fraud detection has long been a focal point of research and is expected to remain an intriguing area of study in the foreseeable future. This enduring interest is primarily driven by the dynamic nature of fraud pat- terns. In this paper, we propose an innovative credit card fraud detection system capable of identifying four distinct patterns of fraudulent transactions using algorithms best suited for each pattern, while also addressing challenges identified by previous researchers in this field. By integrat- ing predictive analytics and an API module for real-time fraud detection, end-users are promptly notified through a graphical user interface (GUI) upon detection of a suspi- cious transaction, empowering fraud investigation teams to take immediate action. We meticulously selected optimal algorithms for each of the four main fraud types through literature review, experimentation, and parameter tuning, as outlined in the methodology. Additionally, we evalu- ated sampling methods to effectively handle the skewed distribution of data.

Our findings underscore the significant impact of re- sampling techniques in achieving higher classifier perfor- mance. The machine learning models identified as LR, NB, LR, and SVM demonstrated the highest accuracy rates in capturing the four fraud patterns (Risky MCC, Unknown web address, ISO Response Code, Transaction above 10000), achieving accuracy rates of 74, 83, 72, and 91 percent, re- spectively. While these models exhibit satisfactory accu- racy levels, our focus moving forward is on enhancing pre- diction capabilities to achieve even better accuracy. More-
over, future extensions of this research aim to explore location- based fraud detection methods.

REFERENCES

[1]. Adi Saputra1, Suharjito2L: Fraud Detection using Machine Learning in eCommerce, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 9, 2019.

[2]. Kaithekuzhical Leena Kurien, Dr. Ajeet Chikkaman- nur: Detection And Prediction Of Credit Card Fraud Transactions Using Machine Learning, International Journal Of Engineering Sciences Research Technology.

[3]. Yashvi Jain, Namrata Tiwari, Shripriya Dubey, Sarika Jain: A Comparative Analysis of Various Credit Card Fraud Detection Techniques, International Journal of Recent Technology and Engineering (IJRTE) 3878, Volume-7 Issue-5S2, January 2019.

[4]. Roy, Abhimanyu, et al: Deep learning detecting fraud in credit card transactions, 2018 Systems and Information Engineering Design Symposium (SIEDS), IEEE, 2018.

[5]. Heta Naik , Prashasti Kanikar: Credit card Fraud Detection based on Machine Learning Algorithms, International Journal of Computer Applications (0975 – 8887) Volume 182 – No. 44, March 2019.

[6]. Navanshu Khare, Saad Yunus Sait: Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models, International Jour- nal of Pure and Applied Mathematics Volume 118 No. 20 2018, 825-838 ISSN: 1314-3395.

[7]. Randula Koralage, , Faculty of Information Tech- nology, University of Moratuwa, Data Mining Techniques for Credit Card Fraud Detection.

[8]. N. Shirodkar, P. Mandrekar, R. S. Mandrekar, R. Sakhalkar, K. M. Chaman Kumar, and S. Aswale, “Credit card fraud detection techniques – A survey,” in 2020 International Conference on Emerging Trends in Information Technology and Engineering (icETITE), 2020, pp. 1–7.

[9]. X. Kewei, B. Peng, Y. Jiang, and T. Lu, “A hybrid deep learning model for online fraud detection,” in 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), 2021, pp. 431–434.

[10]. Suma, V., and Shavige Malleshwara Hills. quot; Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics.quot; Journal of Soft Computing Paradigm (JSCP) 2, no. 02 (2020): 101-110