

# Implementation of Credit Card Fraud Detection

Prof Dipali Dube, Abhay Nighot, Omkar Fulsundar, Prasad Naykodi, Siddhesh Kharge

Department of Computer Engineering

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 machine learning algorithms on an actual dataset and suggest an ensemble-based approach that harnesses the strengths of multiple models. Our experimental outcomes demonstrate the effectiveness of machine learning in accurately identifying fraudulent transactions while minimizing false positives.*

**Keywords:** fraudulent transactions

## I. INTRODUCTION

In the digital era, the widespread adoption of credit cards for financial transactions has transformed the way we handle payments and conduct business. While this convenience 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 substantial financial risks to financial institutions and consumers alike. Detecting and preventing credit card fraud is crucial in today's financial landscape, necessitating a proactive and adaptable approach capable of swiftly identifying fraudulent transactions while minimizing disruptions to legitimate 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 methods and underscoring the pivotal role of machine learning in advancing this field. We delve into the intricacies of data preprocessing, highlighting the significance of adequately preparing datasets to extract meaningful insights.

Additionally, we outline our methodology, encompassing the selection of machine learning algorithms, feature engineering, and the development of an ensemble-based approach that amalgamates the strengths of multiple models. Our experiments and results showcase the practical efficacy of these techniques. We assess our models using various performance metrics, including accuracy, precision, recall, F1-score, and ROC curves, offering a comprehensive 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 unsupervised approaches to hybrid ones. Consequently, understanding the technologies associated with credit card fraud detection and gaining insights into various fraud types have become imperative. Over time, the evolution of fraud patterns has introduced novel forms of fraudulent activities, thereby piquing the interest of researchers. The subsequent 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 incorporation 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 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 fraudulent ones, and vice versa, as pointed out. Handling categorical data is another hurdle, as most machine learning algorithms do not support this type of data. Finally, choosing appropriate detection algorithms and selecting relevant features are cited as challenges in detecting frauds, as training these algorithms can be time-intensive compared to prediction tasks.

### **III. EXPERIMENTAL METHODOLOGY**

#### **Data Overview**

This dataset was compiled by merging two primary sources: the fraud transactions log file and the comprehensive transactions 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 containing fraud cases comprised 200 records, whereas the transaction log file contained 917,781 records.

#### **Data Preparation**

Initially, the raw data were segmented into four distinct datasets based on identified fraud patterns. This segmentation 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. Transactions originating from unknown web addresses. These four datasets were then utilized in two distinct approaches:

**Type A:** Transforming the raw data into numerical representations.

**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, integration, and reduction of data. In the case of Type A preparation, all of these steps were applied to the first three datasets to prepare them numerically. However, for categorical 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. However, 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 removed from the files as they do not contribute to generating meaningful data and could potentially bias the dataset. Furthermore, additional changes were made, including removing 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 illustrates 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. Consequently, data transformation involved normalization, which scales the attribute data to fit within a smaller numeric range.

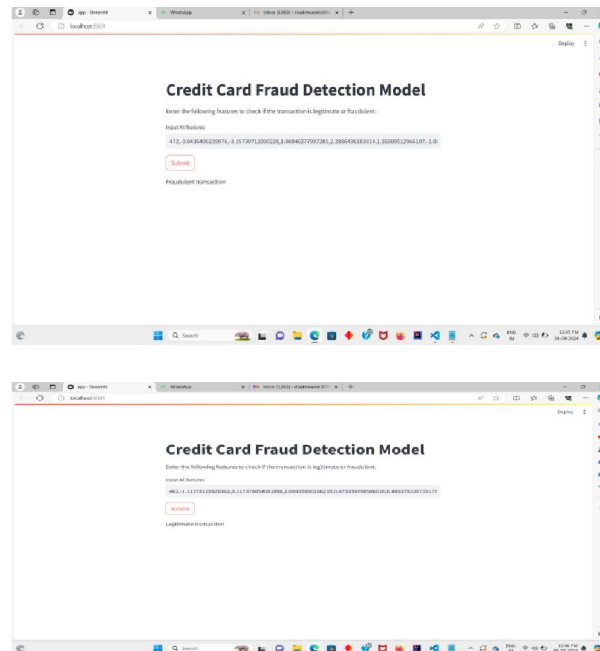
**Data Reduction:**

The chosen strategy for data reduction is dimensionality reduction. This approach aims to mitigate the risk of learning incorrect data patterns, ensuring that the selected features effectively eliminate irrelevant aspects and characteristics of the fraud domain, as highlighted in . Principal Component Analysis (PCA) is a widely recognized method for dimensionality reduction. By applying PCA, the feature 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 patterns. 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 integrating 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 suspicious 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 evaluated sampling methods to effectively handle the skewed distribution of data. Our findings underscore the significant impact of resampling techniques in achieving higher classifier performance. 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, respectively. While these models exhibit satisfactory accuracy levels, our focus moving forward is on enhancing prediction capabilities to achieve even better accuracy. Moreover, future extensions of this research aim to explore locationbased fraud detection methods.

**V. IMPLEMENTATION**



**REFERENCES**

[1]. Adi Saputra1, Suharjo2L: 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 Chikkamannur: 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 Journal of Pure and Applied Mathematics Volume 118 No. 20 2018, 825-838 ISSN: 1314-3395.
- [7]. Randula Koralage, , Faculty of Information Technology, 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. “Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics.” Journal of Soft Computing Paradigm (JSCP) 2, no. 02 (2020):101-110