International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Credit Card Fraud Detection System Using Machine Learning

Alok Kumar Srivastava[1], Anjali Rao[2],Ankita Sahani[3],Shikha Rao[4]
[1]Assistant Professor, Department of Computer Science & Engineering, Buddha Institute of Technology, Gorakhpur, India

[2,3,4]B.Tech Student, Department of Information Technology, Buddha Institute of Technology, Gorakhpur, India

Email- alok6123@gmail.com, anjalirao5apr@gmail.com, ankitasahani329@gmail.com, shikharao23oct@gmail.com

Abstract

Events involving credit card fraud happen regularly and end up costing a lot of money. The number of\son line transactions have grown in large quantities and online. A significant portion of these transactions are credit card transactions. As a result, banks and other financial organizations provide very valuable and in-demand credit card fraud detection programs. Fraudulent transactions can take many different forms and fall under several categories. The four primary fraud incidents in real-world transactions are the topic of this research. A variety of machine learning models are used to address each fraud and the optimal approach is ultimately chosen after examination. This assessment offers a thorough guidance for choosing the best algorithm with regard to the type of frauds and with the use of an appropriate performance measure, we illustrate the evaluation. Real-time credit card fraud detection is another crucial topic that we cover in our project. To determine if a specific transaction is legitimate or fraudulent, we use predictive analytics performed by the implemented machine learning models and an API module. We also evaluate a cutting-edge approach that successfully tackles the skewed distribution of data. According to a private disclosure agreement, the financial institution provides the data for our studies.

Keywords- Skewed distribution, real-time credit card fraud detection, confidential disclosure agreement, fraud detection system, and credit card frauds.

Introduction

With the development of cutting-edge technology and global communication, fraud has been sharply rising. There are two basic strategies to avoid fraud: Credit Card Fraud Detection Based on Transaction Behavior -by John Richard D. Kho, Larry A. They ask the victim to log their personal information in order to fix an issue, and they appear to be from companies like Pay Pal banks, AOL, and eBay. By stealing the victims' identities and later their money, the fraudster can profit. A significant financial loss resulted from credit card fraud.A2017. [1] Both detection and prevention. By serving as an additional layer of defense, prevention thwarts any attacks from fraudsters. Once prevention has failed, detection takes place. Determination hence aids in spotting and warning as soon as a fraudulent transaction is initiated. Web payment gateways are increasingly using card not- present transactions for credit card operations. Online payment systems generated more than $31 trillion in revenue globally in2015, up 7.3% from 2014, according to the Nilson Report from October 2016.Credit card fraud losses globally increased to $21 billion in 2015 and may reach $31 billion by 2020. Although, there has been a sharp rise in fraudulent transactions, which has a significant impact on the

economy. There are various subcategories of credit card fraud. Card-not-present (CNP) and Card-present (CP) frauds are the two primary fraud categories that can be found in a collection of transactions. These two categories can be further divided into behavioral fraud,
Application fraud,theft/counterfeitfraud,andbankruptcyfraud.Ourresearchfocusesonfourtypes of fraud that fall under the CNP fraud category mentioned above, and we offer a way to identify such frauds are al-time fraud. This generation's replacement for such techniques is machine learning, which can handle massive data sets that are difficult for humans to handle. Supervised learning and unsupervised learning are the two primary divisions of machine learning techniques. Fraud detection can be carried out in either way, and the dataset will choose when to employ it. In order to learn under supervision, anomalies must first be classified. Many supervised algorithms have been used to the detection of credit card fraud over the past few years. The two primary analyses of the data employed in this study are categorical analysis and numerical analysis. Initial data in the dataset are categorical.
By using data cleaning and other fundamental preparation methods, the raw data can be prepared. In order to do the evaluation, the relevant procedures must first be used to convert categorical data into numerical data. Second, to discover the best algorithm, machine learning approaches use categorical data. There are often two main criticisms of data mining-based fraud detection research: the dearth of publicly available real data to perform experiments on; and the lack of published well- researched methods and techniques. To counter both of them, this paper garners all related literature for categorization and comparison, selects some innovative methods and techniques for discussion; and points toward other data sources as possible alternative.[2]

Methodology

First of all, we obtained our dataset from Kaggle, a data analysis website which provides datasets. Inside this dataset, there are 31 columns out of which 28 are named as v1-v28 to protect sensitive data. The other columns represent Time, Amount and Class. Time shows the time gap between the first transaction and the following one. Amount is the amount of money transacted. Class 0 represents a valid transaction and 1 represents a fraudulent one. [3] Collecting and sorting raw data, which is then used to train the model to predict the probability of fraud? Machine learning models can recognize unusual credit card transaction and fraud. The first and for emostinvoves Logistic Regression, Random Forest, Naïve Bayes and Multi layer Perception. This chapter discusses the methodology adopted in this study to classify then on-fraudulent transactions from the fraudulent transactions.Figure1 shows the steps used in this work. However, before we discuss the different steps of the methodology used in this work, we first discussed the dataset.

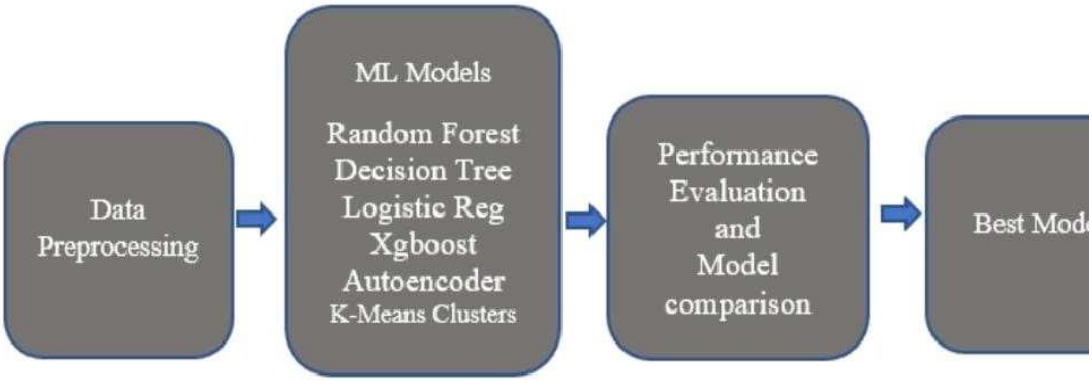

Fig1: Classification Methodology

Literature Review

For the purpose of detecting fraud, a variety of supervised and semi-supervised machine learning techniques are used. However, our goal is to address three key issues with the card fraud dataset, namely, the strong class imbalance, the inclusion of labeled and unlabelled samples, and the need to process a large volume of transactions. To identify fraudulent transactions in real-time datasets, a variety of supervised machine learning algorithms including Decision Trees, Naïve Bayes Classification, Least Squares Regression, Logistic Regression, and SVM are used. Two methods under random forests are used to train the behavioral features of normal and abnormal transactions. These are CART- based and Random- tree- based random forests. A similar research domain was presented by Wen-Fang YU and Na Wang where they used out lie mining, Outlier detection mining and Distance sum algorithms to accurately predict fraudulent transaction in an emulation experiment of credit card transaction dataset of one certain commercial bank. Outlier mining is a field of data mining which is basically used in monetary and internet fields. It deals with detecting objects that are detached from the main system i.e. the transactions that aren't genuine. [4] They have taken attributes of customer's behavior and based on the value of those attributes they've calculated that distance between the observed value of that attribute and its predetermined value Random forest produces good results on tiny sets of data, however there are still some issues when the data is unbalanced. The goal of the upcoming effort will be to solve the aforementioned issue. It is necessary to enhance the random forest method itself. Research is being done on investing at inmate- classifiers and meta-learning methodologies in managing highly skewed credit card fraud data in order to analyze the performance of Logistic Regression, K-Nearest Neighbor, and Nave Bayes. Using supervised learning techniques to identify fraud cases may not always be successful. A deep auto-encoder and restricted Boltzmann machine (RBM) model that may create typical transactions to identify abnormalities in typically occurring patterns. Moreover, a hybrid technique that combines the A ad boost and Majority Voting procedures has been advised.

Fraud is defined as an illegal or criminal deception meant to produce a monetary or personal profit. It is a purposeful action that violates a rule, legislation, or policy with the intention of obtaining unrecognized financial profit. There is a wealth of publicly accessible material on the topic of anomaly or fraud detection in this field that has already been published. Data mining applications, automated fraud detection, and adversarial detection are among the strategies used in this field, according to a thorough review undertaken by Clifton Phua and his coworkers. Suman, Research Scholar, GJUS&T at Hisar HCE, offered methods like supervised and unsupervised learning for credit card fraud detection in a different publication. Although some of these methods and algorithms achieved surprising success, they were unable to offer a reliable, long-term solution to fraud detection.

Future Scope

While we couldn't reach out goal of 100% accuracy in fraud detection, we did end up creating a system that can, with enough time and data, get very close to that goal. As with any such project, there is some room for improvement here. The very nature of this project allows for multiple algorithms to be integrated together as modules and their results can be combined to increase the accuracy of the final result. This model can further be improved with the addition of more algorithms into it. However, the output of these algorithms needs to be in the same format as the others. Once that condition is satisfied, the modules are easy to add as done in the code. This provides a great degree of modularity and versatility to the project. More room for improvement can be found in the dataset.[7] As demonstrated before, the precision of the algorithms increases when the size of dataset is increased. Hence, more data will surely make the model more accurate in detecting frauds and reduce the number of false positives. However, this requires official support from the banks themselves.

Conclusion

Credit card fraud is without a doubt a act of criminal dishonesty. This article has listed out the most common methods of fraud along with their detection methods and reviewed recent findings in this field. Along with the algorithm, pseudo code, explanation of how it is implemented, and results of experimentation, this paper has also provided a detailed explanation of how machine earning can be used to improve fraud detection.

The algorithm does achieve over 99.6% accuracy, but when only a tenth of the data set is considered, its precision is still only 28%. The precision increases to 33% when the algorithm is fed the entire dataset, though. This high accuracy rate is expected given the vast disparity between the number of transactions that are valid and those that are genuine. [5] Complex networks and data mining models share more characteristics than what we could have expected in the first naive approach, most notably having similar objectives: both aim at extra citing information from (potentially complex) systems to ultimately generate new compact quantifiable representations. [5] Credit card fraud is without a doubt an act of criminal dishonesty. This article has listed out the most common methods of fraud along with their detection methods and reviewed recent findings in this field. This paper has also explained in detail, how machine learning can be applied to get better results in fraud detection along with the algorithm, pseudo code, explanation its implementation and experiment action results. The algorithm does achieve over 99.6% accuracy, however when only a tenth of the dataset is considered, its precision is still only 28%. The precision increases to 33% when the system is fed the whole dataset, though. This high accuracy rate is expected given the vast disparity between the number of transactions that are valid and those that are genuine.

Result

The code prints out the number of false positives it detected and compares it with the actual values. This is used to calculate the accuracy score and precision of the algorithms. The fraction of data we used for faster testing is 10% of the entire dataset. The complete data set is also used at the end and both the results are printed. [6]

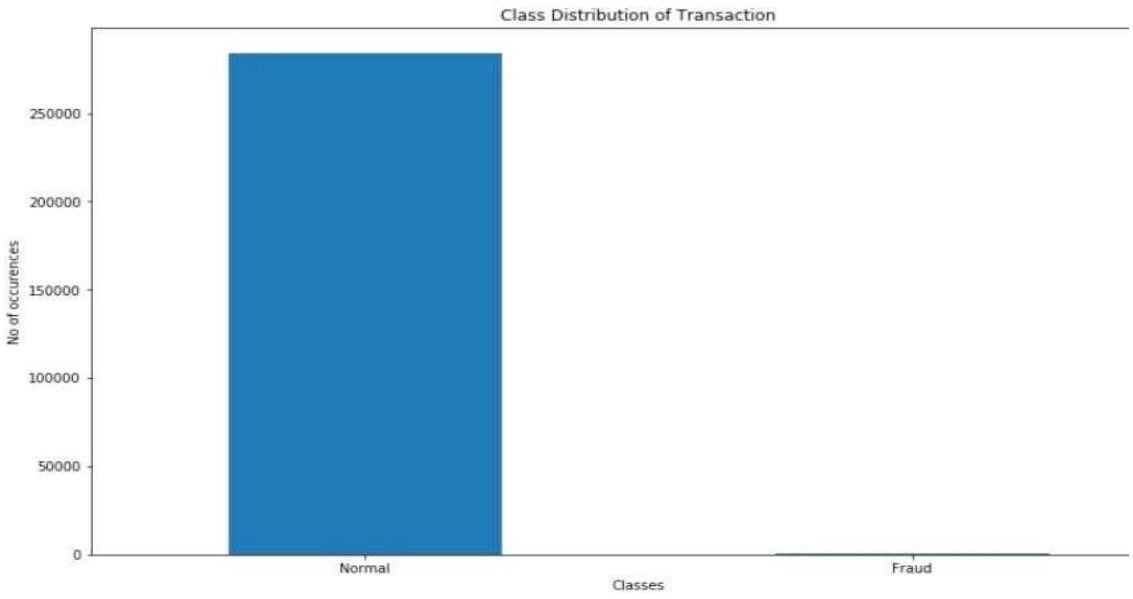

Fig2: Class Distribution Transaction

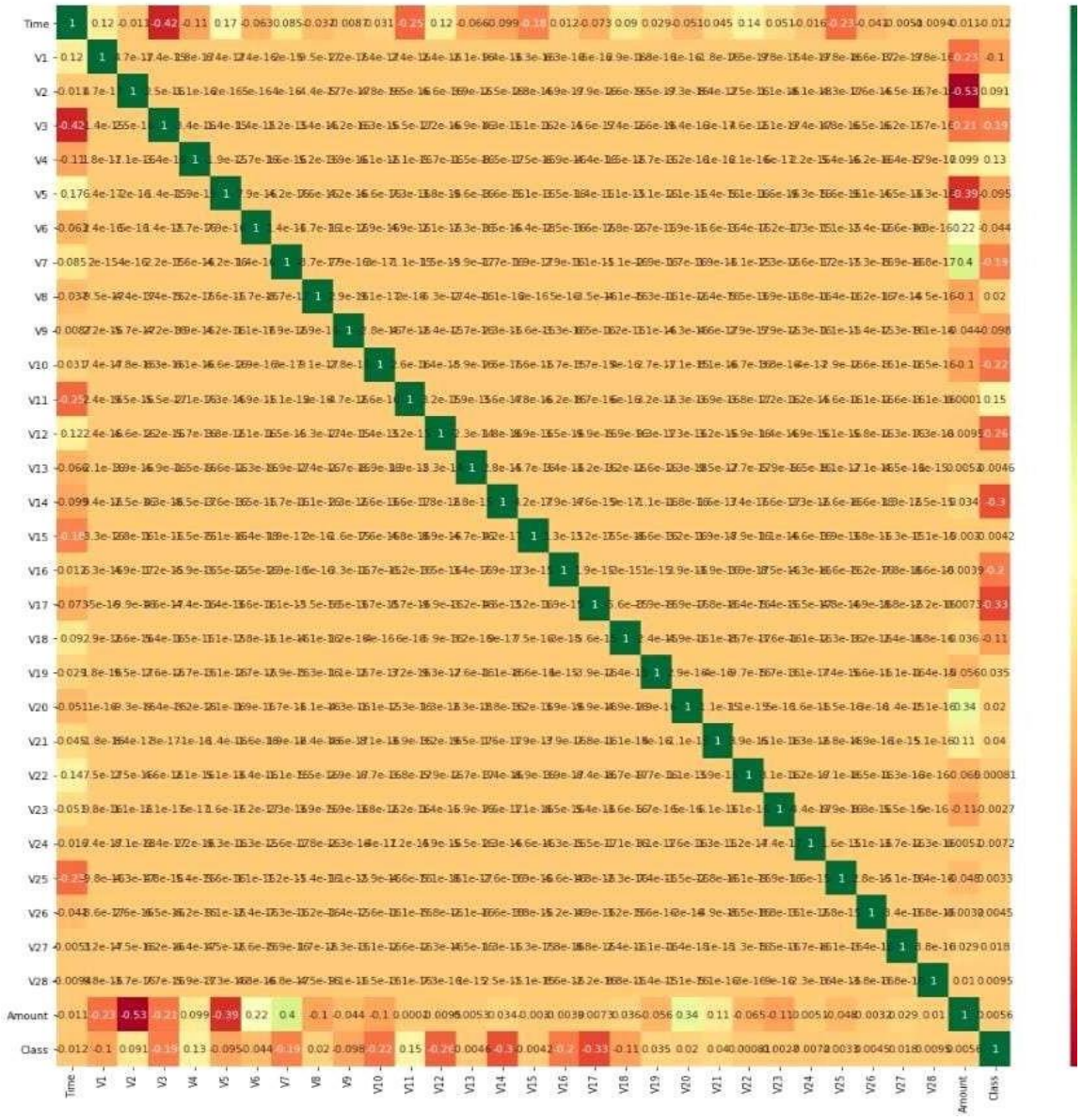

Fig3: Anomaly Detection

References

1. "Credit Card Fraud Detection Based on Transaction Behavior-by John Richard D.Kho, Larry A.Vea" was included in the proceedings of the 2017 IEEE Region10

   Conference (TENCON), which was held in Malaysia from November 5-8, 2017. [1]
2. "A Comprehensive Review of Data Mining-based Fraud Detection Research" by Clifton Phua, Vincent Lee, Kate Smith, and Ross Gayler was published by the School of Business Systems, Faculty of Information Technology, Monash University, Wellington Road, Clayton, Victoria 3800, Australia.
3. Research Scholar, GJUS&T Hisar HCE, Sonepat, "Survey Paper on Credit Card Fraud Detection by Suman," published in International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 3 Issue 3,March2014.
4. Wen-Fang YU and Na Wang's "Study on Credit Card Fraud Detection Model Based on Distance Sum" was published by the 2009 International Joint Conference on Artificial Intelligence.
5. Santiago Moral, Regino Criado, Miguel Romance, and Massimiliano Zanin, "Credit Card Fraud Detection via Parenclitic Network Analysis, "Hindawi Complexity Volume 2018, Article ID 5764370,9 pages.
6. IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 29, NO. 8, AUGUST 2018 "Credit Card Fraud Detection: A Realistic Modeling and a New Learning Method"
7. "Credit Card Fraud Detection-by Ishu Trivedi, Monika, Mrigya, and Mridushi" appeared in the January 2016 issue of the International Journal of Advanced Research in Computer and Communication Engineering.