IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.106108

"CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING TECHNIQUE"

ABHYUDAI A. GORE[1], MAKARAND S. PATRIKAR[2], PRATIK S. KARE[3], VARUN H. JOSHI[4], PROF. A. K. SHAHADE[5]

Dept. of Information Technology Shri. Sant Gajanan Maharaj College of Engineering, Shegaon, Maharashtra, India1-4

Assistant Professor Dept. of IT Shri. Sant Gajanan Maharaj College of Engineering, Shegaon, Maharashtra, India[5]

ABSTRACT: This study primarily focused on detecting credit card fraud in the real world. For the qualified data set, we must first collect credit card data sets. Then, based on the user's responses, deliver inquiries to test the data set, use a credit card. Following the random forest algorithm employing a classification approach with a data set that has previously been examined and supplying a current data set. Finally, the data accuracy of the outcomes is improved. After then, a number of attributes will be processed so that fraud detection can be noticed when looking at the graphical model's depiction. Credit Card Fraud Detection is a typical sample of classification. In this process, we have focused on analyzing and pre-processing data sets as well as the deployment of multiple anomaly detection algorithms such as Local Outlier Factor and Isolation Forest algorithm on the PCA transformed Credit Card Transaction data.

1. INTRODUCTION

Risk assessment is utilised by banks all over the world. Risk rates are analysed using a number of methodologies since credit risk assessment is so critical. Banks divide their customers into groups based on certain characteristics. Client's financial histories and subjective consumer considerations are assessed throughout the exam Those figures are objective, which reflect the financial statements of the company. Fraud detection includes observing and analyzing the behavior of various users in order to estimate the detection of undesired behaviour. We need to understand the many technologies, algorithms, and types involved in detecting credit card fraud in order to identify it efficiently. There are several algorithms for detecting credit card fraud, each with its own set of benefits and accuracy. The methods include:- K-nearest neighbour, Linear regression, Ada Boost, Naive Bayes, J48, Logistic Regression, Random Forest algorithm etc. The null hypothesis is that the credit card transaction is legitimate. As a result, whether it is a correct and real transaction is a false positive, and hence the system model is predicted to be a fraudulent transaction, and a warning is issued. This signifies regular clients looking to make a purchase would defer faraway from making purchases. False negative is a major problem since the transaction is fraudulent but the system model indicates that it is not. A false negative is significantly more serious in our instance than a false positive, because our system model would be costly if it predicted fraudulent transactions as legitimate.

2. LITERATURE SURVEY

I. ACTUAL SYSTEM

In a review of a contextual investigation, which included the identification of Credit Card misrepresentation where information standardization is applied prior to cluster analysis, as well as results obtained from the use of Cluster Analysis and Artificial Neural Network on the discovery of extortion, has revealed that neuronal data sources may be limited by bundling properties in the existing system.Moreover, good results can be obtained by utilising standardised data, which should be MLP prepared. This exam required independent study. The accuracy of estimates is about 50%. The discovery of an estimate and the reduction of the cost measure were noteworthy aspects of this work. The result was 23%, and the formula they came up with was the Bayesian minimal chance[3] . A collective replacement comparison metric that quantifies gains and losses owing to fraud detection is proposed in this system. A cost-sensitive technique based on the Bayes minimum risk is utilised with the existing cost measure.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.106108

II. PLANNED SYSTEM

To classify the credit card data set, we employ the random forest technique in the suggested system. Random Forest is a Regression and Classification algorithm. Irregular words have an advantage over the choice tree in that they adjust the likelihood of overfitting to their set of preparations. To prepare each individual tree, a random subset of the preparation set is evaluated, and each node at that time components of an element are picked from a random subset of the whole list of capabilities, and then a decision tree is formed. In any case, because each tree is generated independently of the others, it is remarkably fast to prepare for large information collections with numerous highlights and information events in random forests[4]. Random Forest ranks the value of variables in a natural way in a regression or classification task. Function class is that the conditional classification target class takes value 1 for positive (fraud) cases and value 0 for negative (non-fraud) cases.

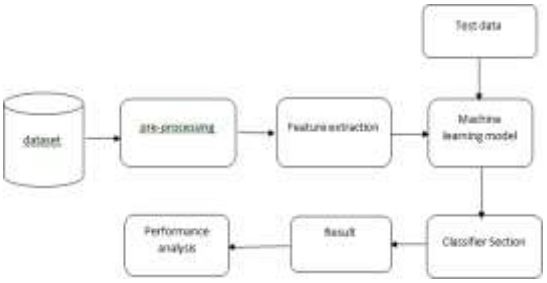


Figure 1: Architecture Diagram

Fraud detection requires a transaction dataset, as well as methods for discovering and classifying fraud. There are several algorithms for detecting fraudulent transactions, therefore start with a literature review to locate some superior methods. Also, improved python algorithms for classifying fraudulent and non- fraudulent transactions should be implemented.

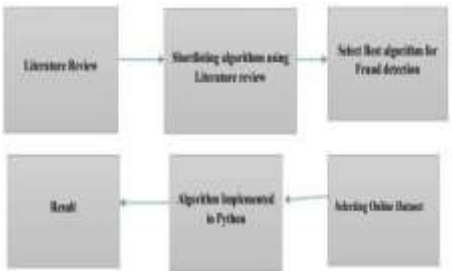


Figure 2: Organization Of Project

This is the path to creating the algorithm. After doing the literature review, we pick up the algorithms to implement and try to recreate them as closely as we can as a python executable. We can then verify the results as we proceed.

III.METHODOLOGY

Various strategies for constructing models based on artificial intelligence, data mining, fuzzy logic, and machine learning have been implemented in the minds of researchers to detect fraudulent activity in credit card transactions. Within fake sample data sets, apps are installed. These data points include the customer's name, age, and account amount, as well as the credit card's origin. So, in the case of card fraud, if the use of cards to commit fraud is higher, the fraud of a credit card transaction will be higher as well, but if this is lower, the level of contribution will be equal. Machine Learning is used to detect and prevent credit card fraud by employing classification and regression techniques. To detect online or offline fraud card transactions, we employ supervised machine learning techniques like Random Forest Algorithms. [5].

IV.ALGORITHM

A. RANDOM FOREST ALGORITHM

Random Forest algorithm is a machine learning based algorithm that combines multiple decision trees together for obtaining efficient outcome. Decision trees are created by random forest algorithm based on data samples and selects the best solution by means of voting. [6]

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.106108

These data points include the customer's name, age, and account amount, as well as the credit card's origin. So, in the case of card fraud, if the use of cards to commit fraud is higher, the fraud of a credit card transaction will be higher as well, but if this is lower, the level of contribution will be equal. Machine Learning is used to detect and prevent credit card fraud by employing classification and regression techniques. To detect online or offline fraud card transactions, we employ supervised machine learning techniques like Random Forest Algorithms. [5].

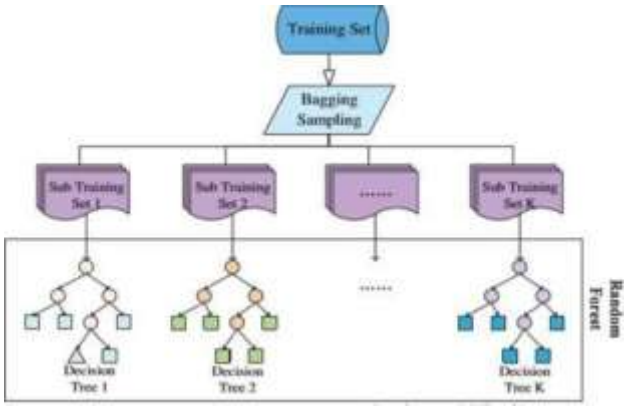


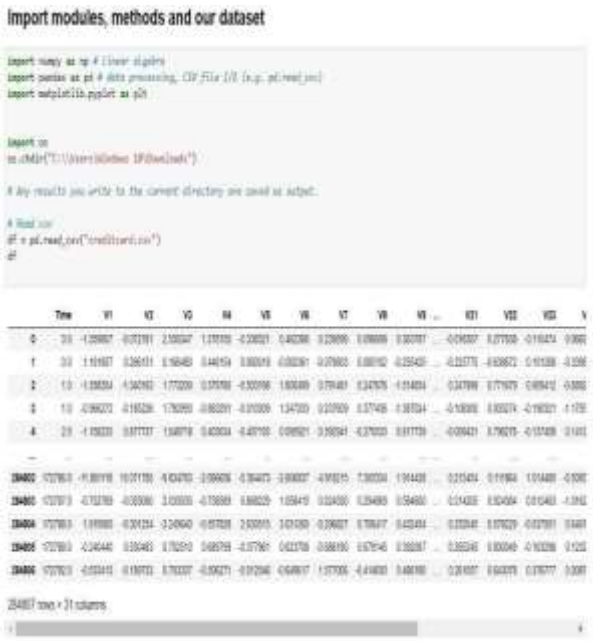
Figure 3: Random Forest Algorithm

The following steps will help you understand how the Random Forest algorithm works.

- Step 1: Begin by selecting randomly selected samples from a dataset.
- Step 2: For each sample, this algorithm will create a decision tree. The forecast result from each decision tree will then be obtained.
- Stage 3: Voting will be done for each expected outcome in this step.
- Step 4: Finally, choose the prediction result with the most votes as the final prediction result

V. MODULES IDENTIFIED

a. DATA COLLECTION



The initial problem starts with collecting the data from various source, enough that we can split the train and test data to get the best fitted model.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.106108

b. DATA PREPROCESSING

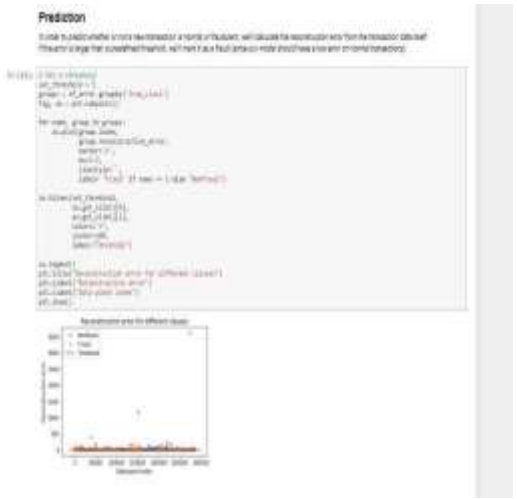
After collecting the data, we need to get rid of incomplete data, data that is polarized and clean the data.

• Formatting: It's possible that you didn't choose the details in a manner that works best for you. The information could be in a proprietary file format and you want it in an electronic database or folder, or it could be in an electronic database and you want it in a spreadsheet.
• Data cleaning: Data cleaning is the process of removing or restoring unfinished or empty data. There may also be data occurrences that are incomplete and do not contain the information you believe you require. You may need to delete these occurrences. Furthermore, there are several properties that contain sensitive information and are likely to be withheld. info that hasn't been completed or that isn't complete. There could also be gaps in the information. Instances of data that do not contain the information that is expected If you want to lever, you'll have to get rid of these happenings. There are other traits that bear weight. Sensitive data, and that the characteristics are likely to be

c.CODE FOR RANDOM FOREST

VI. RESULT ANALYSIS

The algorithm calculates the number of false positives and compares them to genuine data sets. This is used to assess the algorithms' precision and accuracy[10]. The data portion we selected for speedier, more efficient testing was 10% of the total data set. At the end, the entire set of data is used, and both reports are written. These results are included in the output, along with a classification report for each algorithm, where class 0 indicates that the transaction is valid and class 1 indicates that it is fraudulent. For false positives, this result matched the category values.



IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.106108

CONCLUSION

Credit card fraud is a dishonest act that is punishable by law. This article examines new findings in this area, as well as the most typical types of fraud and associated detection methods and algorithms. This research also detailed how machine learning can be used in conjunction with the random forest technique, often known as pseudo code, to improve fraud detection outcomes. While the algorithm achieves above 60% accuracy, it only achieves 28% precision when only a tenth of the data set is considered [9]. However, when the whole data set is entered into the system, the precision rises to 33%.

VII. FUTURE ENHANCEMENTS

Although we did not achieve our aim of 100 percent accuracy in detecting fraud, we were able to develop a programmed with adequate time and resources to achieve something close to it. Because of the project's design, several algorithms can be used as modules, and the outputs of these algorithms can be combined to improve the accuracy of the final result. [12]

VIII.REFERENCES

1. https://www.ijitee.org/wpcontent/uploads/papers/v8i12S/L102 810812
S19.pdf.
2. https://www.researchgate.net/publication/336800562_Credit_Card _Fraud_Detection_using_Machine_Learning_and_Data_Science
3. https://ieeexplore.ieee.org/document/8717766
4. .https://pdfs.semanticscholar.org/6f4a/a57eb9335f6e2658c78a7 a2264e 779a09307.pdf
5. http://www.ijesrt.com/issues%20pdf%20file/Archive- 2019/March2019/26.pdf
6. https://www.ijrte.org/wpcontent/uploads/papers/v7i6s4/F1044 0476S4 19.pdf
7. A. Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams, and P.Beling, "Deep learning detecting fraud in credit card transactions," in 2018 Systems and Information Engineering Design Symposium (SIEDS). IEEE, 2018, pp.
129–134.
8. D. Dighe, S. Patil, and S. Kokate, "Detection of credit card fraud transactions using machine learning algorithms and neural networks: A comparative study," in 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). IEEE, 2018, pp. 1–6.
9. M. Puh and L. Brkic´, "Detecting credit card fraud using selected machine learning algorithms," in 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE, 2019, pp. 1250–1255.
10. J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," in 2017 International Conference on Computing Networking and Informatics (ICCNI). IEEE, 2017, pp. 1–9.
[11] V. Dheepa and R. Dhanapal, "Behavior based credit card fraud detection using support vector machines," ICTACT Journal on Soft computing, vol. 6956, pp. 391–397, 2012.