ISSN (O) 2321-2004, ISSN (P) 2321-5526

IJIREEICE

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 9, Issue 12, December 2021

DOI: 10.17148/IJIREEICE.2021.91209

Credit Card Fraud Detection Using Machine Learning

Mrs. Prof. Lakshmipraba Balaji[1], Mr. Janardhan umale [2],Ms. Prajakta Tembare [3], Mr. Prasnna Kerutagi[4]

Professor Department of Electronics and Telecommunication,

Dr. D.Y. Patil Institute of Engineering Management and Research Akurdi, Pune, Maharashtra, India[1]

Students Department of Electronics and Telecommunication,

Dr. D.Y. Patil Institute of Engineering Management and Research Akurdi, Pune, Maharashtra, India[2,3,4]

Abstract: Credit card fraud is the most serious problem in today's world, and there is an urgent need to combat it. "Credit card fraud is the process of cleaning dirty money, making the source of funds untraceable." On a daily basis, huge amounts of money are exchanged in the global market, making detecting credit card fraud activity a difficult task. As previously stated, (Anti-credit card fraud Suite) is introduced to detect suspicious activities, but it is only applicable to individual transactions and not to other bank account transactions. To address these issues, we propose a machine learning method based on 'Structural Similarity.' This method identifies common attributes and behaviour with other bank account transactions. It is difficult to detect credit card fraud transactions from large datasets, so we propose case reduction methods to reduce the input dataset and then find pairs of transactions with other bank accounts that share common attributes and behaviour.

Keywords: Machine learning ,SVM algorithm

I. INTRODUCTION

Every year, credit card fraud consumes up to 5% of the world's GDP (Gross Domestic Product). The goal of using AI to combat credit card fraud is to detect suspicious activity. To combat credit card fraud, most entities that complete financial transactions must keep detailed records of their clients' accounts and activities. They are required to report any information that appears to be suspicious to the government for further investigation. If suspicious data is detected, the transaction records are checked to detect credit card fraud activity. In this case, we use Artificial Intelligence and Machine Learning Algorithms to detect suspicious activities and solve them by training on the data associated with those activities. We will employ both supervised and unsupervised algorithm techniques.

Card fraud is classified into two types: card-present fraud (which is less common nowadays) and card-not-present fraud (more common). The compromise can happen in a variety of ways,

and it usually happens without the cardholder's knowledge. Because of the internet, database security lapses have become especially costly; in some cases, millions of accounts have been compromised. Cardholders can report stolen cards quickly, but the details of a compromised account may be held by a fraudster for months before any theft, making it difficult to identify the source of the compromise. The cardholder may not be aware of fraudulent use until he or she receives a statement. Cardholders can reduce their risk of fraud by frequently checking their accounts to ensure there are no suspicious or unknown transactions. When a credit card is lost or stolen, it can be used for illegal purchases until the holder notifies the issuing bank and the account is blocked. To encourage prompt reporting, most banks provide free 24-hour telephone numbers. Even so, a thief may be able to make unauthorised purchases on a card before it is cancelled.

Types of Frauds:

- Online and Offline
- Card Theft
- Data phishing
- Application Fraud
- Telecommunication Fraud

II. BACKGROUND AND RELATED WORK

Fraud in any way is a criminal activity and is an offence, credit card fraud is stealing money. There are many studies in which they tried to find whether a transaction is fraud or not. Still having many challenges and tries to overcome those problems Firstly, many used Data Mining Techniques to find fraudulent transactions by using some Traditional approach, which is not conventional and these days fraudsters are so smart that they can do fraud without violating rules so, Using

Vol. 9, Issue 12, December 2021
DOI: 10.17148/IJIREEICE.2021.91209

Machine learning is conventional. Machine learning also comes with challenges. So, here a heavily imbalanced data set is considered, so that it will give us the best algorithm to use along with its challenges. As it is heavy imbalanced, even if the proposed algorithm is good or not, it gives us an accuracy of about 99.9%. So, here the under sampling is considered to provide us with good results as in, Outlier detection and removal algorithms are used to accurately predict fraudulent transactions of a credit card transaction dataset as in Outlier data is used to deal with detecting the anomalous activities. Even after a huge number of proposed algorithms and mechanis ms to stop fraudulent transactions, the fraudsters are so clever that they always try to find new ways to make anonymous transactions and sometimes even the proposed algorithm could not find whether the transaction is fraud or not. So, to stop these frauds, the proposed algorithm should be made to learn from the past frauds and use it for future frauds, which can even detect the fraud before it takes place

III. PROPOSED METHODOLOGY

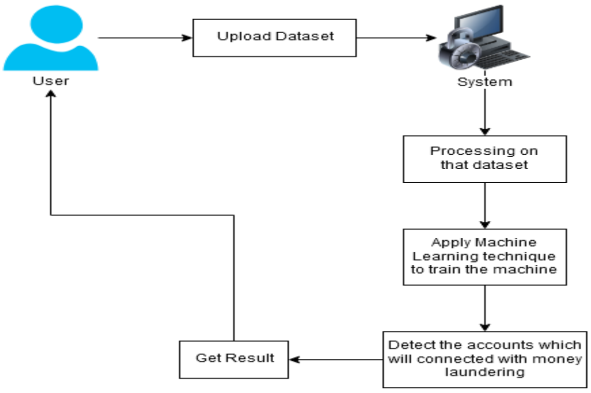
User
Upload Dataset
System
Processing on that dataset
Apply Machine Learning technique to train the machine
Detect the accounts which will connected with money laundering
Get Result

Figure 1.System Diagram

Using SVM Classifier

The Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for classification as well as regression. Though we call it a regression problem, it is best suited for classification. The SVM algorithm's goal is to find a hyperplane in an N-dimensional space that clearly classifies the data points. The size of the hyperplane is determined by the number of features. If there are only two input features, the hyperplane is simply a line. When the number of input features reaches three, the hyperplane transforms into a two-dimensional plane. When the number of features exceeds three, it becomes difficult to imagine. Support vector machines (SVMs) are supervised machine learning algorithms that are both powerful and flexible. They are used for classification and regression. However, they are most commonly used in classification problems. SVMs were first introduced in the 1960s, but they were refined in 1990. When compared to other machine learning algorithms, SVMs have a distinct implementation method. They have recently become extremely popular due to their ability to handle multiple continuous and categorical variables.

SVM Operation

An SVM model is essentially a representation of various classes in a multidimensional hyperplane. The hyperplane will be generated iteratively by SVM in order to minimise error. SVM's goal is to divide datasets into classes in order to find the maximum marginal hyperplane. SVMs are classified into two types, each of which is used for a different purpose

Simple SVM: This type of SVM is commonly used for linear regression and classification problems.

Kernel SVM: Has more flexibility fornon-linear data because it can fit a hyperplane rather than a two-dimensional space.

Why are SVMs used in machine learning?

SVMs are used in a variety of applications, including handwriting recognition, intrusion detection, face detection, email classification, gene classification, and web page generation. One of the reasons we use SVMs in machine learning is for this reason. It can perform classification and regression on both linear and non-linear data. Another reason we use SVMs is that they can discover complex relationships between your data without requiring you to perform numerous transformations on your own. It's an excellent choice when working with smaller datasets with tens to hundreds of thousands of features. Because of their ability to handle small, complex datasets, they typically produce more accurate results than other algorithms.

Logistic Regression: Logistic Regression is one of the classification algorithm, used to predict a binary values in a given set of independent variables (1 / 0, Yes / No, True / False). To represent binary / categorical values, dummy variables are used. For the purpose of special case in the logistic regression is a linear regression, when the resulting variable is categorical then the log of odds are used for dependent variable and also it predicts the probability of occurrence of an event by fitting data to a logistic function.

Random Forest:Random forest is a tree based algorithm which involves building several trees and combining with the output to improve generalization ability of the model. This method of combining trees is known as an ensemble method. Ensembling is nothing but a combination of weak learners (individual trees) to produce a strong learner. Random Forest can be used to solve regression and classification problems. In regression problems, the dependent variable is continuous. In classification problems, the dependent variable is categorical.

Naive Bayes: a similar method to predict the probability of different class based on various attributes. This algorithm is mostly used in text classification and with problems having multiple classes.

Algorithm steps:

Step 1: Read the DATASET.

Step 2: Random Sampling is done on the data set to make it balanced.
Step 3: Divide the dataset into two parts i.e., Train dataset and Test dataset.
Step 4: Feature selection are applied for the proposed models.
Step 5: Accuracy and performance metrics has been calculated to know the efficiency for different algorithms.
Step6: Then retrieve the best algorithm based on efficiency for the given dat

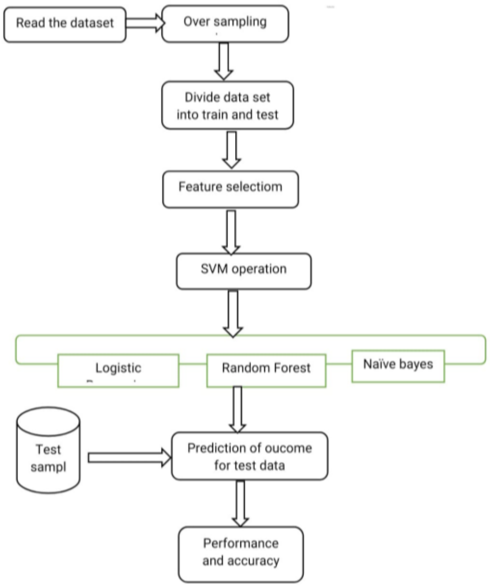

Figure 2. Workflow overview. (depicts workflow and steps of our implementation to reach the objective.)

IV. FUTURE SCOPE:

From the above analysis, it is clear that many machine learning algorithms are used to detect the fraud but we can observe that the results are not satisfactory. So, we would like to implement deep learning algorithms to detect credit card fraud accurately.

In future we can provide more security using different techniques.

Also in future we can work on android application also for better efficiency.

V. CONCLUSION

The proposed ML framework aims to find potential Credit card fraud groups among a large number of financial transactions. In order to improve the efficiency of the framework, case reduction methods such as matching transaction detection and balance score filter are used to narrow down the list of potential ML accounts. Next by taking advantage of structural similarity, we can identify and group potential Credit card accounts. Our preliminary experimental results show a high degree of accuracy in detection of ML account.

VI.REFERENCES:

[1]S. Bhattacharyya, S. Jha, K. Tharakunnel, J. C. Westland, "Data mining for credit card fraud: A comparative study", Decision Support Systems, vol. 50, no. 3, pp. 602-613, 2011.

DOI: 10.17148/IJIREEICE.2021.91209

[2]Y. Sahin, E. Duman, "Detecting credit card fraud by ANN and logistic regression", Innovations in Intelligent Systems and Applications (INISTA) 2011 International Symposium, pp. 315-319, 2011.

[3]Selvani Deepthi Kavila,LAKSHMI S.V.S.S.,RAJESH B " Automated Essay Scoring using Feature Extraction Method " IJCER ,volume 7,issue 4(L), Page No. 12161-12165. [15] S.V.S.S.Lakshmi,K.S.Dee

[4]A Novel Approach for Credit Card Fraud Detection using Decision Tree and Random Forest Algorithms [2021]

[5]Yashvi Jain, Namrata Tiwari, ShripriyaDubey, Sarika Jain, "A Comparative Analysis of Various Credit Card Fraud Detection Techniques, Blue Eyes Intelligence Engineering and Sciences Publications 2019"

[6] Learning Robert A. Sowah, Moses A. Agebure, Godfrey A. Mills, Koudjo M. Kaumudi, "New Cluster Undersampling Technique for Class Imbalance "of 2016 IJMLC

[7]Baraneetharan, E. "Role of Machine Learning Algorithms Intrusion Detection in WSNs: A Survey." Journal of Information Technology 2, no. 03 (2020): 161-173

[8] Mitra, Ayushi. "Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset)." Journal of Ubiquitous Computing and Communication Technologies (UCCT) 2, no. 03 (2020): 145-152.