IJCRT.ORG   **ISSN : 2320-2882**

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Review: Detecting Frauds in Credit Card Using Best Machine Learning Techniques

Rucha Narkhede [1] , Nilesh Chaudhari [2]

PG Student [1] , Assistant Professor [2]

Department of Computer Engineering,

Godavari College of Engineering, Jalgaon, India

Abstract: Now-a-days online transactions have become essential component of our lives. Also financial fraud is an ever growing menace with far results in financial industry, government and corporate organizations. Fraud can be defined as criminal deception with intent of learning financial gain. Credit card frauds are comfortably targets by fraudsters. As e-commerce and several other online sites have enhanced the online payment transactions, increases the risk for online frauds. Credit card fraud generally happens when the card was stolen for any of the unauthorized aims, or even when the fraudster utilizes the credit card information for his use. Therefore in the introduces world, we're experiencing a lots of credit card problems. To decrease fake transactions, machine acquiring algorithmic program savor Naive Bayes, Logistic Regression, J48, AdaBoost and Random Forest, etc. are discussed. Similar set of Algorithmic program are implemented and tested utilizing an online dataset. The work implemented in Java. The performance based on the accuracy and low false positive rate. Over comparative study, it has the ability to be concluded that Naive Bayes and Random Forest algorithmic program are compose greater in credit card fraud detection in spite of other algorithm.

Index Terms - Data Mining, Machine Learning, Credit Card Fraud Transactions, classification

I. INTRODUCTION

Because of increasing use of E-Commerce, over there has been big use of credit cards for online activities which led to a big number of frauds narrated to credit cards. The most important goal is to make a fraud detection algorithm, which finds the fraud transactions with less time and lofty accuracy by utilizing machine acquiring based classification algorithms. As tech is moving send rapidly, the payment by cash is reduced and online payment receives increased, this facilitates way for the fraudsters to make unidentified transactions. [2].

There're a lot of two types of credit card frauds. One is theft of physical card, and other one is stealing sensitive information from the card, as an example card number, cvv code, type of card and other. By stealing credit card information, a fraudster can negate a big number of money or make a big amount of buy prior to cardholder finds out. As a consequence of that, firms use several machine acquiring methods to peer which transactions are fake and which are not. The objective of this paper is to analyze several machine learning algorithms, for instance Logistic Regression (LR), Random Forest (RF), Naive Bayes (NB) to work out which algorithm is the vast majority of proper for credit card fraud detection.[3].

Fraud detection entails monitoring and analyzing the behavior of several users in order to estimate detect or avoid undesirable behavior. In order to identify credit card fraud detection productively, we need to understand the various machine learning technologies, algorithms and types involved in detection of credit card frauds.

II. RELATED WORK

Deep learning is part of the main and considerable strategies being used for the detection of fraud within the credit card. These kinds of networks have a posh distribution of knowledge that is incredulous troublesome to acknowledge. Deep auto encoders are used in several stages to extract the vast majority of efficient features of the info and for classification functions. Also, higher accuracy and low variance are accomplished amongst these networks[4].

The research on credit card fraud detection utilizes both Machine Learning and Deep Learning algorithms. In this section, we improve the task done in two distinct points : (i) The methods that are readily accessible for fraud detection, and (ii) The strategies that are accessible to the imbalanced information in the dataset. To handle the imbalanced information several of the strategies are available. They're as (1) classification methods (2) sampling methods (3) resembling techniques. In this place are several of the Machine Learning algorithms that are utilized for credit fraud detection are support vector machine (SVM), decision trees, logistic regression, gradient boosting, k-nearest neighbor, etc[1].

Supervised learning categorizes the dataset utilizing training information whereas unsupervised learning categorizes it utilizing clustering technique. A current near was to solve this plight utilizing deep learning as said by AutoEncoder and Confined Boltzmann Machine. The writers concluded that supervised learning is greater proper for historical database in credit card fraud detection. The set aside of Undersampling on the posterior probability of machine learning version is analyzed by utilizing Bayesian minimum risk theory. Support Vector Machines besides   with Random Forests are discussed . The entity utilized in paper utilizes decision tree which is price tag sensitive. The paper uses data mining technology to create credit card acquiring fraud analysis version as said by mass credit card transaction information   and merchant materials, and additionally   developed merchant fraud risk management system[5].

In this paper, discuss about the imbalance of class and how to handle it and additionally discuss how to work on big dataset. Detection of credit card fraud for new frauds will be problematic if new information has severe changes in fraud patterns. Replacing the version is hazardous as machine learning algorithm take big time for training in spite of predicting[6].

Classification of credit card transactions is mostly a binary classification problem. Here, credit card transaction is or as a legitimate transaction (negative class) or a fake transaction (positive class). Fraud detection is basically viewed as a data mining classification problem, where the purpose is to appropriately categorize the credit card transactions as legitimate or fraudulent[8].

III. SYSTEM ARCHITECTURE

An overview of the complete process of credit card fraud detection is shown in the diagram in figure 1, each of whose steps are explained in this section. The preliminary step for the detection of fraud and non-fraud cases is data collection and necessary preprocessing to convert it into a form, which can be used by the learning algorithms.
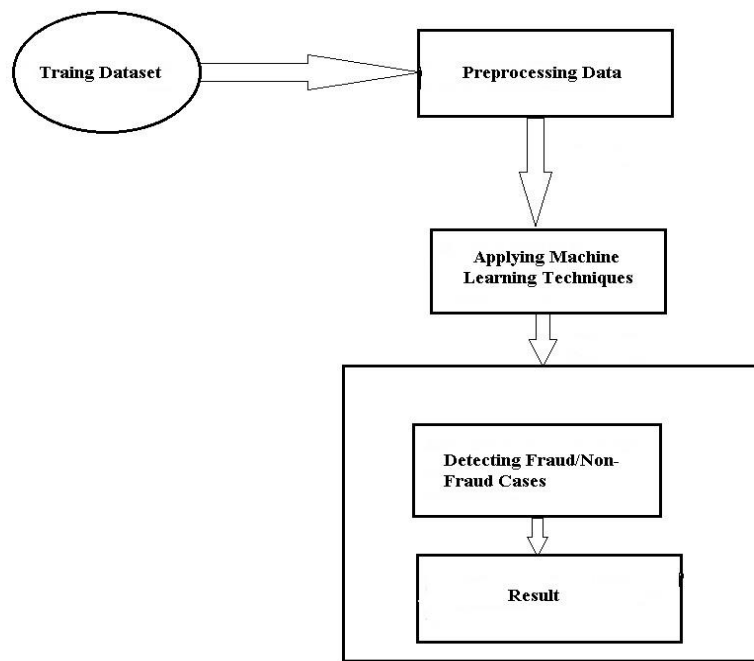
figure 1

IV. PROPOSED METHODOLOGY

On the Literature review, many algorithms are applied on credit card fraud detection. On the survey basis Naive Bayes, Logistic Regression, J48, AdaBoost and Random Forest are better than the other algorithms for credit card fraud detection.

A. Naive Bayes

Naive Bayes algorithm is a supervised learning algorithm, which is based on Bayes theory. It is very simple and powerful algorithm and it is used for classification problems. It is primarily used for classification of text, that encompasses high-dimensional dataset. Naive Bayes Classifier helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts the probability of an object. It is utilized to decide the probability of a speculation with prior knowledge. It depends on the conditional probability.

P (A/B) = (P (B/A) P (A)) / P (B)

Where, **P(A) is Prior Probability**: Probability of hypothesis before observing the evidence.

P(B) is Marginal Probability: Probability of Evidence

P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.

P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

☐ Naive Bayes algorithm is very easy and fast.

B. Logistic Regression

Logistic regression is part of the vast majority of machine learning algorithms, which comes below the Supervised Learning Method. It is utilized for predicting the categorical dependent variable utilizing a delivered set of self-reliant variables. Logistic regression predicts the output of a categorical dependent variable. Because the result need to be a categorical or discrete value. It is the ability to be or accurate or false, 0 or 1, Yes or No, etc. but rather than that of delivering the correct rate value as 0 and 1, it delivers the probabilistic values which is between 0 and 1.

☐ This algorithm to same as linear regression algorithm. But linear regression is utilized for solving Regression issues and logistic regression is utilized for solving classification problem.
☐ Linear regression categorized as:
☐ **Binomial** – 2 Conceivable types (i.e. 0 or 1) only
☐ **Multinomial** – 3 or more Conceivable types and which are not ordered
☐ **Ordinal** – Ordered in kind (i.e. extremely poor, destitute , good, extremely good)
☐ This algorithm simple for binary and multivariate classification task[7].

C. J48

J48 is a machine learning decision tree classification algorithm that is built on the top of the Iterative Dichotomiser 3. It helps evaluates the data categorically and continuously. This concept essentially refers to a multivariate decision tree where instances are classified by multiple attributes at each internal node.

☐ J48 algorithmic program utilized to produce a decision tree and it's for classification task.
☐ J48 is an extended of ID3 (Iterative Dichotomiser 3). This algorithm has several exceptional options such as, rules derivation, continues assessment range, decision tree pruning, etc.
☐ The J48 algorithm is utilized to categorized distinct applications and compose correct outcomes of the classification. J48 algorithm is part of the finest machine learning algorithm to contemplate the information categorically and continuously.

☐ This algorithm is working on fixed and categorical variables[7].

D. AdaBoost

☐ AdaBoost is the acronym for Adaptive Boosting which is machine learning technique used as an Ensemble Method. The most widely used algorithm with AdaBoost is decision trees with one level. This implies that the decision in the AdaBoost has only a single split.

☐ For AdaBoost, Each instance in the training dataset is weighted. Initial weight is set to: **Weight (xi) =   (1/n)**
Where, **xi – ith training instance,**
 n – Number of training instance
☐ This algorithm mainly used for classification rather than regression. So that AdaBoost algorithm is used in fraud detection because this classify the transaction which transactions are frauds and which are non-frauds[7].

E. Random Forest

The Random Forest algorithm is part of one of the most significant supervised learning algorithm. This algorithm is utilized for both regression and classification purposes. But, this algorithm is widely utilized for classification problems. Generally, a forest is furious up of trees and similarly, the Random Forest algorithm made the decision trees on the sample information and receive the prediction from each of the sample data. Later Random Forest algorithm is a co-ordinates method. This algorithm is greater than the single decision trees due to it reduces the over-fitting by equating the result[1].

Steps for Random Forest Algorithm :

1. Take the dataset from Kaggle for credit card fraud that is trained and randomly choose several of the sample data.

2. Utilizing the randomly generated sample data made the Decision Trees that are utilize to categorise the cases into the fraud and non-fraud cases.

3. The Decision Trees are formed by splitting the nodes, the nodes which have the highest Information acquire make it as the root node and categorize the fraud and non-fraud cases.

4. Now the majority vote is performed and the decision trees outcome if 0 then it is non-fraud cases and if the result 1 then it is fraud cases.

5. Finally, we locate the accuracy, precision, recall, and F1 -score for both the fraud and non-fraud cases.

V. CONCLUSION

In credit card fraud detection, we frequently deals with highly imbalanced datasets. For the chosen dataset from Kaggle, we will show that our proposed algorithms are able to detect fraud transactions with very lofty accuracy and low false positive rate. Hence for better performance, our result shows that classification of algorithms done by preprocessing data rather than raw data. Because of forwarding preprocessing data, output of algorithms is with lofty accuracy and give best results.

REFERENCES

[1] R. Sailusha, V. Gnaneswar, R. Ramesh and G. R. Rao, "Credit Card Fraud Detection Using Machine Learning," *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2020, pp. 1264-1270, 2020

[2] D. Tanouz, R. R. Subramanian, D. Eswar, G. V. P. Reddy, A. R. Kumar and C. V. N. M. Praneeth, "Credit Card Fraud Detection Using Machine Learning," *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 967-972, 2021.

[3] D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic and A. Anderla, "Credit Card Fraud Detection - Machine Learning methods," *2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)*, pp. 1-5, 2019.

[4] https://www.jetir.org/papers/JETIR2204420.

[5] D. Dighe, S. Patil and S. Kokate, "Detection of Credit Card Fraud Transactions Using Machine Learning Algorithms and Neural Networks: A Comparative Study," *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 2018, pp. 1-6.

[6] Varun Kumar K S , Vijaya Kumar V G , Vijayshankar A , Pratibha K, 2020, Credit Card Fraud Detection using Machine Learning Algorithms, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 07 (July 2020).

[7] https://www.ijcaonline.org/archives/volume182/number44/naik-2019-ijca-918521.pdf

[8] J. O. Awoyemi, A. O. Adetunmbi and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," *2017 International Conference on Computing Networking and Informatics (ICCNI)*, 2017, pp. 1-9.