ENGINEERING IN ADVANCED RESEARCH SCIENCE AND TECHNOLOGY

ISSN 2350-0174
Vol.03, Issue.02
December-2022
Pages: 1185-1189

Credit card Fraud Detection Using Machine Learning

Dammu Karunamma [1], Punuri Sahaja [2], Kondasingu roopanjali [3], Shaik Abesha[4]

[1]*M. Tech., Dept of CSE, Qis college of Engineering and technology JNTUK University. AP, India karuna.d@qiscet.edu.in*

[2]*B. Tech., Dept of CSE, Qis college of Engineering and technology , JNTUK University. AP, India sahajapunuri@gmail.com*

[3]***B. Tech., Dept of CSE,** Qis college of Engineering and technology **JNTUK University. AP, India rooparoopanjali7842@gmail.com***

[4]***B. Tech., Dept of CSE,** Qis college of Engineering and technology **JNTUK University. AP, India abeshask2001@gmail.com***

Abstract:- Credit card fraud detection is presently the most frequently occurring problem in the present world. This is due to the rise in both online transactions and e-commerce platforms. Credit card fraud generally happens when the card was stolen for any of the unauthorized purposes or even when the fraudster uses the credit card information for his use. In the present world, we are facing a lot of credit card problems. To detect the fraudulent activities the credit card fraud etection system was introduced. This project aims to focus mainly on machine learning algorithms. The algorithms used are random forest algorithm and the Ada boost algorithm. The results of the two algorithms are based on accuracy, precision, recall , and F1-score. The ROC curve is plotted based on the confusion matrix. The Random Forest and the Ada boost algorithms are compared and the algorithm that has the greatest accuracy, precision, recall , and F1-score is considered as the best algorithm that is used to
detect the fraud.

Keywords-Block chain, credit card fraud, fraudulent activities, Random Forest, Adaboost, ROC curve

I. INTRODUCTION:-

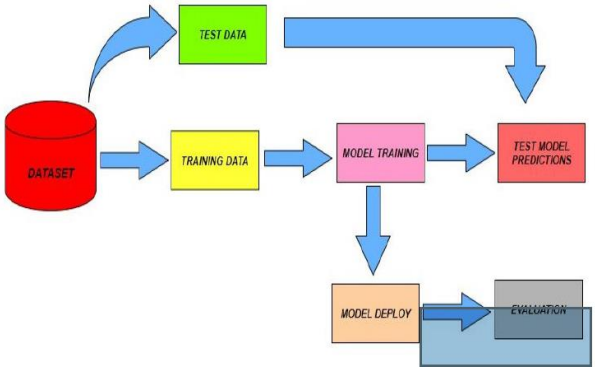
The online shopping growing day to day. Credit cards are used for purchasing goods and services with the help of **virtual card** and **physical card** whereas virtual card for online transaction and physical card for offline transaction. In a physical-card based purchase, the cardholder presents his card physically to a merchant for making a payment. To carry out fraudulent transactions in this kind of purchase, an attacker has to steal the credit card. If the cardholder does not realize the loss of card, it can lead to a substantial financial loss to the credit card company. In online payment mode, attackers need only little information for doing fraudulent transaction (secure code, card number, expiration date etc.). In this purchase method, mainly transactions will be done through Internet or telephone. To commit fraud in these types of purchases, a fraudster simply needs to know the card details. Most of the time, the genuine cardholder is not aware that someone else has seen or stolen his card information. The only way to detect this kind of fraud is to **analyse the spending patterns** on every card and to figure out any inconsistency with respect to the "**usual" spending patterns.** Fraud detection based on the analysis of existing purchase data of cardholder is a promising way to reduce the rate of successful credit card frauds. Every cardholder can be represented by a set of patterns containing information about the typical purchase category, the time since the last purchase, the amount of money spent, etc. Deviation from such patterns is a potential threat to the system.

Banks collect a lot of historical records corresponding to millions of customer's transactions. They are credit card and debit card operations, but unfortunately, only a small portion, if any, is open access. Fraud detection is a critical problem affecting large financial companies that have increased due to the growth in credit card transactions. The proposed method consists of the Predictive modeling and Logistic Regression. Now a day's bank transactions as well as credit card frauds increased. One of the most target frauds are credit card fraud, the fraud can occur any type of credit products, such products are retail, home loan and personal loan. To detect credit card fraud, data mining techniques-Predictive modeling and Logistic Regression are used. In prediction model to predict the continuous valued functions. Credit card of CSV files will be analyzed to predict the outcome. In this paper, we propose to detect credit card transaction using available data set and data mining techniques of predictive modeling, Decision tree, and Logistic Regression. Predictive modeling splits the data into two partitions 70% of testing and 30%of training check output class distribution to predict the outcome. The decision tree to get the result as a tree with root node describes the best predictor in the data, the combination of two or more branches is denoted by decision node (non leaf nodes) and each branch represents a value for the attribute which is tested. The leaf node may be 1 in the case of fraud and 0 otherwise. Logistic regression or logistic model is a regression model, where the dependent variable is categorical of a linear generalized model.

II. Related Work:- New methods for credit card fraud detection with a lot of research methods and several fraud detection techniques with a special interest in the neural networks, data mining, and distributed data mining. Many other techniques are used to detect such credit card fraud. When done the literature survey on various methods of credit card fraud detection, we can conclude that to detect credit card fraud there are many other approaches in Machine Learning itself. The research on credit card fraud detection uses both Machine Learning[1][2] and Deep Learning algorithms[7]. In this section, we enhance the work done in two different points:(i) the methods that are readily available for fraud detection, and (ii) The techniques

that are available to handle the imbalanced data. To handle the imbalanced data A[11] some of the techniques are available. They are (a) classification methods (b) sampling methods (c) resembling techniques. Here are some of the Machine Learning algorithms that are used for credit fraud detection are support vector machine(SVM), decision trees, logistic regression, gradient boosting, K-nearest neighbor, etc; In 2019, Yashvi Jain, NamrataTiwari, Shripriya Dubey, Sarika jain have researched various techniques[10] for credit cards fraud detection such as support vector machines(SVM), artificial neural networks(ANN), Bayesian Networks, Hidden Markov Model, K-Nearest Neighbors (KNN) Fuzzy Logic system and Decision Trees. In their paper, they have observed that the algorithms k-nearest neighbor, decision trees, and the SVM give a medium level accuracy. The Fuzzy Logic and Logistic Regression give the lowest accuracy among all the other algorithms. Neural Networks, naive bayes, fuzzy systems, and KNN offer a high detention rate. The Logistic Regression, SVM, decision trees offer a high detection rate at the medium level. There are two algorithms namely ANN and the Naïve Bayesian Networks which perform better at all parameters. These are very much expensive to train. There is a major drawback in all the algorithms. The drawback is that these algorithms don't give the same result in all types of environments. They give better results with one type of datasets and poor results with another type of dataset. Algorithms like KNN and SVM give excellent results with small datasets and algorithms like logistic regression and fuzzy logic systems give good accuracy with raw and un sampled data. In 2019, Heta Naik, Prashasti Kanikar, has done their research on various algorithms [4] like Naïve Bayes, Logistic Regression, J48, and Adaboost. Naïve Bayes on among the classification algorithm. This algorithm depends upon Bayes theorem. Bayes's theorem finds the probability of an event that is occurring is given. The Logistic regression algorithm is similar to the linear regression algorithm. The linear regression is used for the prediction or forecasting the values The logistic regression is mostly used for the classification task. The J48 algorithm is used to generate a decision tree and is used for the classification problem. The J48 is the extension of the ID3 (Iterative Dichotomieser). J48 is one of the most widely used and extensively analyzed areas in Machine Learning. This algorithm mainly works on constant and categorical variables. Adaboost is one of the most widely used machine learning algorithms and is mainly developed for binary classification. The algorithm is mainly used to boost the performance of the decision tree. This is also mainly used for the classification of the regression. The Adaboost algorithm is fraud cases to classify the transactions which are fraud and non-fraud. From their work they have concluded that the highest accuracy is obtained for both the Adaboost and Logistic Regression. As they have the same accuracy the time factor is considered to choose the best algorithm. By considering the time factor they concluded that the Adaboost algorithm works well to detect credit card fraud.In 2019 Sahayasakila V, D.Kavya Monisha, Aishwarya, Sikhakolli Venkatavisalakshiswshai Yasaswi have explained the Twain important algorithmic techniques [8] which are the Whale Optimization Techniques (WOA) and SMOTE (Synthetic Minority Oversampling Techniques). They mainly aimed to improve the convergence speed and to solve the data imbalance problem. The class imbalance problem is overcome using the SMOTE technique and the WOA technique. The SMOTE technique discriminates all the transactions which are synthesized are again re-sampled to check the data accuracy and are optimized using the WOA technique. The algorithm also improves the convergence speed, reliability, and efficiency of the system. In 2018 Navanushu Khare and Saad Yunus Sait have explained their work [5] on decision trees, random forest, SVM, and logistic regression. They have taken the highly skewed dataset and worked on such type of dataset . The performance evaluation is based on accuracy, sensitivity, specificity, and precision. The results indicate that the accuracy for the Logistic Regression is 97.7%, for Decision Trees is 95.5%, for Random Forest is 98.6%, for SVM classifier is 97.5%. They have concluded that the Random Forest algorithm has the highest accuracy among the other algorithms and is considered as the best algorithm to detect the fraud. They also concluded that the SVM algorithm has a data imbalance problem and does not give better results to detect credit card fraud.

SYSTEM ARCHITECTURE



III. Literature Survey:-

Credit Card Fraud Detection using Various Methods and Techniques Credit card fraud detection has drawn lot of interest and a number of techniques, with special emphasis on data mining and neural networks, have been proposed to counter fraud in this field. Vatsa et al. [2] determine the effectiveness of neural network for credit card fraud detection. The neural network used for this study is the P-RCE (Restricted Coulomb Energy) neural network. The authors concluded that it was possible to achieve a reduction of 20% to 40% in the total fraud losses. Aleskerov et al. [25] presented CARDWATCH, a database mining system based on a neural network learning module. The system trains a neural network with the past data of a particular customer, which can then be used to process the current spending behavior and detect anomalies and they assume that since the normal behavior of the thief is to purchase as much as possible in limited time, the anomaly in transactions will most probably be detected. Chan et al divide a large data set of transactions into smaller subsets and then apply the mining techniques in parallel in a distributed data mining approach [26]. The resultant base models are then combined

to generate a meta-classifier. More recently, Syeda et al have discussed the use of parallel granular neural networks for fast credit card fraud detection [27]. The Parallel Granular Neural Network (PGNN) aims at speeding up the data mining and knowledge discovery process. Suvasini panoramic et al. [21] suggested a fusion approach to find the suspicion level of the transaction. In this process, a hybrid approach is derived that combines both the supervised and unsupervised learning approaches. This system consist of four components, namely, rule based filter, Dempstershafer adder, transaction history database and Bayesian learner. In the rule based component, they determine the suspicion level of each incoming transaction based on the extent of its deviation from good pattern. Dempster –Shafer's theory is used to combine multiple such evidences and an initial belief is computed. The transactions are classified as normal, abnormal or suspicious depending on this initial belief. Once a transaction is found to be suspicious, belief is further strengthened or weakened according to its similarity with fraudulent or genuine transactions history using Bayesian learning.

IV. Conclusion:- This process is used to detect the credit card transaction, which are fraudulent or genuine. Data mining techniques of Predictive modeling, Decision trees and Logistic Regression are used to predict the fraudulent or genuine credit card transaction. In predictive modeling to detect and check output class distribution. The prediction model predicts continuous valued functions. We have to detect 148 may be fraud and other are genuine. In decision tree generate a tree with root node, decision node and leaf nodes. The leaf node may be 1 becomes fraud and 0 otherwise. Logistic Regression is same as linear regression but interpret curve is different. To generalize the linear regression model, when dependent variable is categorical and analyzes relationship between multiple independent variables. -based framework for implementing insurance transaction processes as smart contracts. Experiments conducted to study the scalability clearly showed the parameters used during block chain creation should be chosen carefully, as they have a direct effect on the network latency. Though the database is currently not encrypted, it can be encrypted with fine-grained access control. In our model, each smart contract has its own set of endorsing peers, and this can be extended even to the transaction level, to enable separate set of endorsing peers for each transaction.

V.REFERENCES

[1] B.M. Ayyub, Elicitation of Expert Opinions for Uncertainty and Risks, CRC Press, Roca Raton, 200 1.
[2] Bandyopadhyay, S., & Wolfe, J. (2000). A critical review of online auction models. Journal of the Academy of Business and Economics.
[3] Barry Scholnick, Nadia Massoud, Anthony Saunders, Santiago Carbo-Valverde Francisco Rodriguez-Fernhdez (2008) The economics of credit cards, debit cards and ATMs: A survey and some new evidence, Journal of Banking & Finance 32 (2008) 1468-1483.
[4] Berger, A. N., Frame, W. S., & Miller, N. (2005). Credit scoring and the availability, price, and risk of small business credit. Journal of Money, Credit, and Banking, 37 (2), 191-222.
[5] Brown, J. D. (1988). Illusion of well-being: A social psychological perspective of mental health. Psychological Bulletin, 103, 193-2 10.
[6] Ausubel, L. (1991). 'The failure of competition in the credit caid market. The American Economic Review, 8 l(1). 50-8 1.
[7] "l'xample of a supplement chargeable to the customer %hen paying by credit card". Bmibaby.com. 201 1-01-05.
[8] Byungtae Lee, Hyungjun Cho, Myungsin Chae, Seonyoung Shim (2010) Empirical analysis of online auction fraud: Credit card phantom transactions, Expert Systems with Applications, 37, 2010, pp.2991- 2199.
[9] Calem, P. S., & Mester, L. J. (1995). Consumer behavior and the stickiness of credit-card interest rates. The American Economic Review, 85(5), 1327- 1336.
[10] Chiun-Sin Lin, Gwo-Hshiung Tzeng, Yang-Chieh Chin (2010) Combined rough set theory and flow network graph to predict customer chum in credit card accounts, Expert Sy.items with Applications, In Press.