

IDENTIFICATION OF CREDIT CARD THEFT USING VARIOUS MACHINE LEARNING

Ms. B.TRIVENI¹, Mr.K.NAGARAJU²

#1 Assistant professor in the department of IT at DVR & DR. HS MIC College of
Technology (Autonomous), Kanchikacherla, NTR District.

#2 MCA student in the Department of Computer Applications (DCA) at DVR & DR.

HS MIC COLLEGE OF TECHNOLOGY, Kanchikacherla, NTR District

ABSTRACT:

In this paper we mainly focus on credit card fraud detection in real world. Here the credit card fraud detection is based on fraudulent transactions. Generally, credit card fraud activities can happen in both online and offline. But in today's world online fraud transaction activities are increasing day by day. So, in order to find the online fraud transactions various methods have been used in existing system. In proposed system we use Random Forest Algorithm (RFA), SVM, DECISION Tree algorithms for finding the fraudulent transactions and the accuracy of those transactions. This algorithm is based on supervised learning algorithm where it uses decision trees for classification of the dataset. After classification of dataset a confusion matrix is obtained. The performance of Random Forest Algorithm is valuated based on the confusion matrix. The results obtained

from processing the dataset gives accuracy of about 90%.

INTRODUCTION

Credit card generally refers to a card that is assigned to the customer (cardholder), usually allowing them to purchase goods and services within credit limit or withdraw cash in advance. Credit card provides the cardholder an advantage of the time, i.e., it provides time for their customers to repay later in a prescribed time, by carrying it to the next billing cycle. Credit card frauds are easy targets. Without any risks, a significant amount can be withdrawn without the owner's knowledge, in a short period. Fraudsters always try to make every fraudulent transaction legitimate, which makes fraud detection very challenging and difficult task to detect. In 2017, there were 1,579 data breaches and nearly 179 million records among which Credit card frauds were the most common form with 133,015 reports,

then employment or tax-related frauds with 82,051 reports, phone frauds with 55,045 reports followed by bank frauds with 50,517 reports from the statistics released by FTC Taxonomy for Frauds. With different frauds mostly credit card frauds, often in the news for the past few years, frauds are in the top of mind for most of the world's population. Credit card dataset is highly imbalanced because there will be more legitimate transactions when compared with a fraudulent one. As advancement, banks are moving to EMV cards, which are smart cards that store their data on integrated circuits rather than on magnetic stripes, have made some on-card payments safer, but still leaving card-not-present frauds on higher rates. According to 2017 the US Payments Forum report, criminals have shifted their focus on activities related to CNP transactions as the security of chip cards were increased. Shows the number of CNP fraud cases that were registered in respective years. Even then there are chances for thieves to misuse the credit cards. There are many machine learning techniques to overcome this problem

Problem Statement:

There are various fraudulent activities detection techniques that have been implemented in credit card transactions have been kept in researcher minds to methods to develop

models based on artificial intelligence, data mining, fuzzy logic and machine learning. Credit card fraud detection is significantly difficult, but also a popular problem to solve. In our proposed system we built the credit card fraud detection using Machine learning. With the advancement of machine learning techniques. Machine learning has been identified as a successful measure for fraud detection. A large amount of data is transferred during online transaction processes, resulting in a binary result: genuine or fraudulent. Within the sample fraudulent datasets, features are constructed. These are data points namely the age and value of the customer account, as well as the origin of the credit card. There are hundreds of features and each contributes, to varying extents, towards the fraud probability.

Motivation:

In existing system, research about a case study involving credit card fraud detection, where data normalization is applied before Cluster Analysis and with results obtained from the use of Cluster Analysis and Artificial Neural Networks on fraud detection has shown that by clustering attributes neuronal inputs can be minimized. And promising results can be obtained by using normalized data and data should be MLP trained. This research was based on

unsupervised learning. Significance of this paper was to find new methods for fraud detection and to increase the accuracy of results. The data set for this paper is based on real life transactional data by a large European company and personal details in data is kept confidential. Accuracy of an algorithm is around 50%. Significance of this paper was to find an algorithm and to reduce the cost measure. The result obtained was by 23% and the algorithm they find was Bayes minimum risk.

RELATED WORK

Detecting Credit Card Fraud by Decision Trees and Support Vector Machines

With the developments in the Information Technology and improvements in the communication channels, fraud is spreading all over the world, resulting in huge financial losses. Though fraud prevention mechanisms such as CHIP&PIN are developed, these mechanisms do not prevent the most common fraud types such as fraudulent credit card usages over virtual POS terminals or mail orders. As a result, fraud detection is the essential tool and probably the best way to stop such fraud types. In this study, classification models based on decision trees

and support vector machines (SVM) are developed and applied on credit card fraud detection problem. This study is one of the firsts to compare the performance of SVM and decision tree methods in credit card fraud detection with a real data set.

BLAST-SSAHA Hybridization for Credit Card Fraud Detection

A phenomenal growth in the number of credit card transactions, especially for online purchases, has recently led to a substantial rise in fraudulent activities. Implementation of efficient fraud detection systems has thus become imperative for all credit card issuing banks to minimize their losses. In real life, fraudulent transactions are interspersed with genuine transactions and simple pattern matching is not often sufficient to detect them accurately. Thus, there is a need for combining both anomaly detection as well as misuse detection techniques. In this paper, we propose to use two-stage sequence alignment in which a profile analyser (PA) first determines the similarity of an incoming sequence of transactions on a given credit card with the genuine cardholder's past spending sequences. The unusual transactions traced by the profile analyser are next passed on to a deviation analyser (DA)

for possible alignment with past fraudulent behaviour. The final decision about the nature of a transaction is taken on the basis of the observations by these two analysers. In order to achieve online response time for both PA and DA, we suggest a new approach for combining two sequence alignment algorithms BLAST and SSAHA.

Credit Risk Assessment in Commercial Banks Based on Support Vector Machines

According to the practical situation of credit risk assessment in commercial banks, a set of index system is established. The index system includes financial indexes and non-financial indexes. Then support vector machines (SVM) algorithm is used for assessment in this research. In the method, training sets are selected by the increasing proportions. Proportions are determined by the number of samples. In order to verify the effectiveness of the method, a real case is given and the experimental results show that the model has high correct classification accuracy.

Credit Risk Assessment for Rural Credit Cooperatives Based on Improved Neural Network

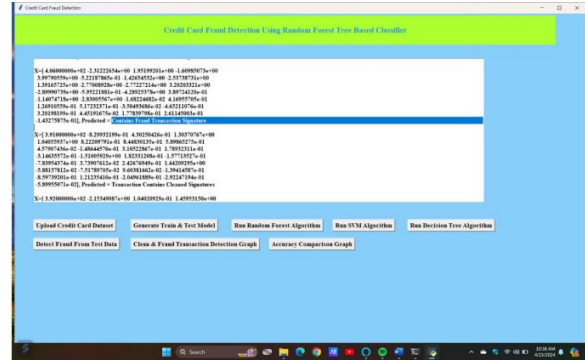
At present, the researches on credit risk analysis mainly focus on commercial bank loan or consumer credit risk, and there is little research about the credit risk of rural credit cooperatives. The purpose of this paper is to evaluate credit risk for the rural credit cooperatives using artificial neural network model. We establish credit risk assessment index system for rural credit cooperatives. Then, a kind of credit risk assessment model based on particle swarm optimized neural network is put forward. The proposed scheme has fast convergence speed and high accuracy. We can conclude that the propose model has good development prospect in the field of credit risk assessment.

METHODOLOGY

1. Upload credit card dataset: we collected the credit card fraud data from Kaggle website.
2. Generate train and test model: we have to preprocess the collected data like cleaning null values, unwanted rows, unwanted columns. After that we have to split the data into two parts training part with 80% and testing part with 20%.
3. Run algorithms: we have to train the training data for RF, SVM, Decision

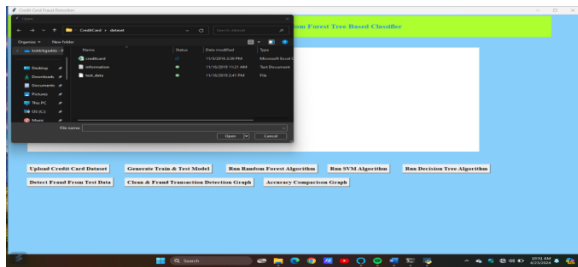
Tree algorithm and test with test data to get accuracy.

4. Detect Credit card Fraud from test data: by using random forest we can detect the fraud signatures.
5. Clean and fraud graph: we can display the clean and fraud transactions with graph.

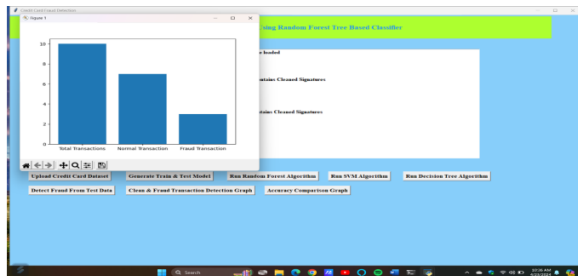


In above Screen is the output of Detect Fraud Test Data

RESULT AND DISCUSSION



In above Screen is the output of upload credit card dataset.



In above Screen is the output of Accuracy Comparison Graph

CONCLUSION

The Random Forest Algorithm will perform better with a larger number of training data, but speed during testing and application will suffer. Application of more pre-processing techniques would also help. The SVM algorithm still suffers from the imbalanced dataset problem and requires more preprocessing to give better results at the results shown by SVM is great but it could have been better if more preprocessing have been done on the data.

REFERENCES

- [1] Shaath G: Credit Risk Analysis and Prediction Modelling of Bank Loans Using R, vol. 8, no-5, pp. 1954-1966.
- [2] LI Changjiang, HU Peng: Credit Risk Assessment for Ural Credit Cooperatives based on Improved Neural Network,

International Conference on Smart Grid and Electrical Automation vol. 60, no. - 3, pp 227-230, 2017.

[3] Wei Sun, Chen-Guang Yang, Jian-Xun Qi: Credit Risk Assessment in Commercial Banks Based on Support Vector Machines, vol.6, pp 2430-2433, 2006.

[4] Amlan Kundu, Suvasini Panigrahi, Shamik Sural, Senior Member, IEEE, “BLASTSSAHA Hybridization for Credit Card Fraud Detection”, vol. 6, no. 4 pp. 309-315, 2009.

[5] Y. Sahin and E. Duman, “Detecting Credit Card Fraud by Decision Trees and Support Vector Machines, Proceedings of International Multi Conference of Engineers and Computer Scientists, vol. I, 2011.

AUTHOR PROFILES:



Ms.B TRIVENI completed her M.TECH. Currently working as an Assistant Professor in the department of IT at DVR & DR HS MIC COLLEGE OF TECHNOLOGY(Autonomous), Kanchikacherla, NTR(Dist), AP. Her areas of

interest are Data Mining, Cloud Computing and Machine Learning & Networks.



Mr.K.NAGARAJU as MCA Student in the Department of DCA at DVR &DR.HS MIC COLLEGE OF TECHNOLOGY, Kanchikacherla, NTR(DT).He completed his BSC(General) in SVL Kranthi Degree College.His areas of interests are C,Java,Python,Web development.