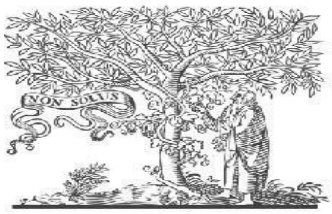
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.com

COPY RIGHT



IJIEMR Transactions, online available on 29th Aug 2021.

Link :http://www.ijiemr.com/downloads.php?vol=Volume-10&issue=ISSUE-08

DOI: 10.48047/IJIEMR/V10/I08/19-6

Title:- **CREDIT CARD FRAUD DETECTION USING RANDOM FOREST AND CART ALGORITHM**

Volume 10, Issue 08, Pages:162-167

Paper Authors
Mrs.T.RajyaLakshmi[1], Bezawada Sravani [2], Shaik Umera Anjum [3] , Thota Teja Veerasai [4] , Dasari Avinesh [5]

Editor IJIEMR

www.ijiemr.com

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.com

CREDIT CARD FRAUD DETECTION USING RANDOM FOREST AND CART ALGORITHM

Mrs.T.RajyaLakshmi[1], Bezawada Sravani [2], Shaik Umera Anjum [3] , Thota Teja Veerasai [4] , Dasari Avinesh [5]
[1]Assistant Professor, Dept. of CSE, [2]17ME1A0571, [3]17ME1A0572, [4]17ME1A0586, [5]17ME1A05A6
Ramachandra College of Engineering, A.P., India

ABSTRACT:

As we know that the usage of credit cards transactions has beenincreased drastically. With the increase in utilization of credit cards, there is a considerable rise in fraudulent activities too. So, the project is mainly focused on credit card transactions in the real world. The objective of the intruders is to obtain the products/goods without crediting the amount from their accounts or by crediting it from other accounts. Previously there are many unsupervised machine learningtechniques like ANN for credit card fraud detection which yields less accuracy. So with the technology advancements, in this project we uses the supervised machine learning techniques like Random forest & Cart algorithms in order to increase the accuracy of the model. Accuracy is considered because the performance of the techniques is evaluated based on accuracy, specificity, sensitivity & precision.

1. INTRODUCTION

There are various fraudulent activities detection techniques has 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 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 transferredduring online transaction processes, resulting in a binary result: genuine or fraudulent. Within the sample fraudulent datasets, featuresare 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.

In existing System, a research about a case study involving creditcard 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 beMLP 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 isbased 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

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.com

they find was Bayes minimum risk.There are various fraudulent activities detection techniques has 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 popular problem to solve. In
our proposed system we built the credit card fraud detection using Machine learning.

2. RELATED WORK

Existing System

In Existing System, Cluster Analysis (K-Means Clustering) & ANNis used to detect the credit card fraud transactions. Before Cluster Analysis the data get normalized and the results obtained after applying Cluster Analysis and ANN takes the minimal neural inputs because theinputs gets reduced in normalization by eliminating the redundant data. The data should be MLP (Multilayer Perceptron) trained in order to get promising results. MLP uses supervised algorithm called Back Propagation (used for improving the accuracy of predictions by using gradient descent which adjusts the parameters in such a way that its output deviation is minimal) for training the data set. These are basedon unsupervised learning with 50% accuracy. So, the main objective of our project is to use the supervised learning algorithms like Random Forest and Cart Algorithms in order to give more accuracy.

Proposed System

In proposed system, Random forest & Cart algorithms are usedto detect the credit card fraud transactions in order to improvise the accuracy of prediction. These algorithms are supervised ML techniques. Random forest is a collection of decision trees which is used for classification and regression. With respect to decision trees, random forest has a advantage that it corrects the habit of overfitting to training sets. From the whole training set, a subset is sampled randomly so that individual tree is trained and then decision trees are built, from here each node in the tree is then splits on a feature selected from randomly sampled subset of the whole feature set. Even though there is a huge data sets Random forest works very fast to train them because
individual tree is trained independently of others.

3. METHODOLOGY:

End User Upload the dataset and when end user had uploaded,the system will in return gives the response that the dataset what the end user have uploaded was successful. In the next step user perform the preprocessing with the help of the system and the system in return give the response to the user. When the user give the test data then based upon the training data by applying the algorithms the system will detect the fraud transaction.

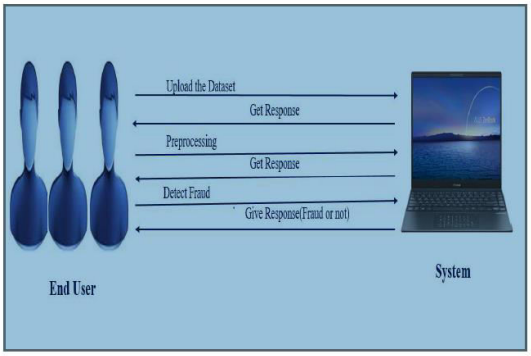


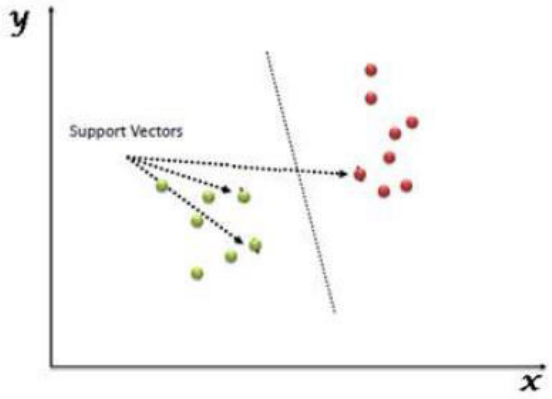
Figure 1: Architecture

Feature selection is an important part in machine learning toreduce data dimensionality and extensive research carried out for a reliable feature selection method. For feature selection filter method and wrapper method have been used. In filter method, features are selected on the

basis of their scores in various statistical tests that measure the relevance of features by their correlation with dependent variable or outcome variable. Wrapper method finds a subset of features by measuring the usefulness of a subset of feature with the dependent variable. Hence filter methods are independent of any machine learning algorithm whereas in wrapper method the best feature subset selected depends on the machine learning algorithm used to train the model. In wrapper method a subset evaluator uses all possible subsets and then uses a classification algorithm to convince classifiers from the features ineach subset. The classifier consider the subset of feature with which the classification algorithm performs the best. To find the subset, the evaluator uses different search techniques like depth first search, random search, breadth first search or hybrid search. The filter method uses an attribute evaluator along with a ranker to rank all the features in the dataset. Here one feature is omitted at a time that has lower ranks and then sees the predictive accuracy of the classification algorithm. Weights or rank put by the ranker algorithms are different than those by the classification algorithm. Wrapper method is useful for machine learning test whereas filter method is suitable for data mining test because data mining has thousands of millions of features.

Algorithms used in this project

Support Vector Machine

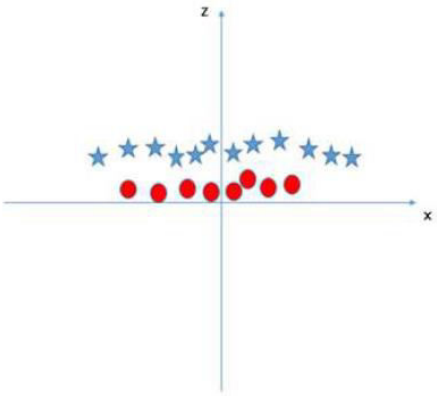


"Support Vector Machine" (SVM) is a supervised machine learningalgorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well (look at the below snapshot). Support Vectors are simply the co-ordinates of individual observation. The SVM classifier is afrontier which best segregates the two classes (hyper-plane/ line).You can look at support vector machines and a few examples of its working here.we got accustomed to the process of segregating the two classes with a hyper-plane. Now the burning question is "How can we identify the right hyper-plane?".Don't worry, it's not as hard as you think!Identify the right hyper-plane (Scenario-1): Here, we have three hyper-planes (A, B and C). Now, identify the right hyper-plane toclassify star and circle.You need to remember a thumb rule to identify the right hyperplane: "Select the hyper-plane which segregates the two classes better". In this scenario, hyper-plane "B" has excellently performed this job. Identify the right hyper-plane (Scenario-2): Here, we have three hyper-

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.com

planes (A, B and C) and all are segregating the classes well.

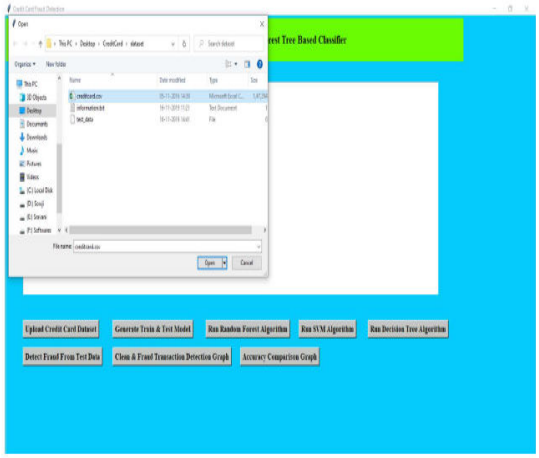
All values for z would be positive always because z is thesquared sum of both x and y In the original plot, red circles appear close to the origin of x and y axes, leading to lower value of z and star relatively away from the origin result to higher value of z. In the SVM classifier, it is easy to have a linear hyper-plane between these two classes. But, another burning question which arises is, should we need to add this feature manually to have a hyper-plane. No, the SVM algorithm has a technique called the kernel trick. The SVM kernel is a function that takes low dimensional input space and transforms it to a higher dimensional space i.e. it converts not separable problem to separable problem. It is mostly useful in non-linear separation problem. Simply put, it does some extremely complex data transformations, then finds out

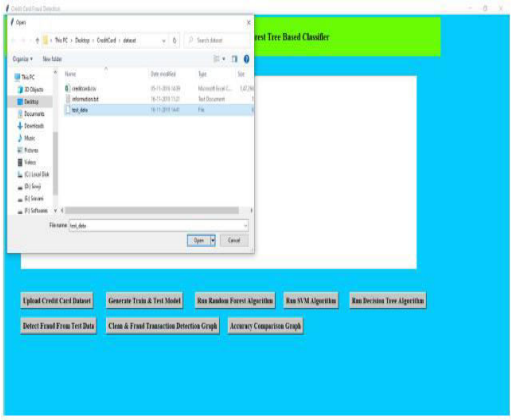


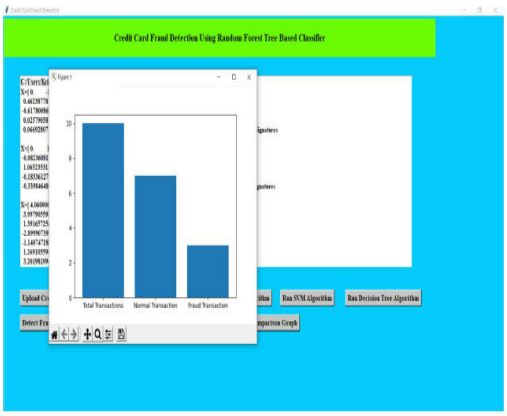
Now, How can we identify the right hyper-plane?Here, maximizing the distances between nearest data pointn(either class) and hyper-plane will help us to decide the right hyperplane.
This distance is called as Margin. Let's look at: Above,Above, you can see that the margin for hyper-plane C is high as compared to both A and B. Hence, we name the right hyper-plane as C. Another lightning reason for selecting the hyper-plane with higher margin is robustness. If we select a hyper-plane having low margin then there is high chance of miss-classification.Identify the right hyper-plane (Scenario-3):Hint: Use the rules as discussed in previous section to identify the right hyper-planeSome of you may have selected the hyper-plane B as it has higher margin compared to A. But, here is the catch, SVM selects the hyper-plane which classifies the classes accurately prior to maximizing margin. Here, hyper-plane B has a classification error and A has classified all correctly. Therefore, the right hyper-plane is A.As I have already mentioned, one star at other end is like an outlier for star class. The SVM algorithm has a feature to ignore outliers and find the hyper-plane that has the maximum margin. Hence, we can say, SVM classification is robust to outliers.

4. STUDY OF RESULTS:

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.com







5.CONCLUSION :

The Random forest algorithm will perform better with a largernumber 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 pre-processing 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.Application of more pre-processing techniques would also help. The SVM algorithm still suffers from the imbalanced dataset problem and requires more pre-processing 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.

6.REFERENCES :

[1] Sudhamathy G: Credit Risk Analysis and PredictionModelling of Bank Loans Using R, vol. 8, no-5, pp. 1954- 1966.

[2] LI Changjian, HU Peng: Credit Risk Assessment for uralCredit Cooperatives based on Improved Neural Network,International Conference on Smart Grid and ElectricalAutomation 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] AmlanKundu, SuvasiniPanigrahi, Shamik Sural, Senior Member, IEEE, "BLAST-SSAHA 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.