

CREDIT CARD FRAUD DETECTION AND BEHAVIORAL PATTERN USING RANDOM FOREST ALGORITHM AND NEURAL NETWORKS

KUNDETI RAMYA, Student[CSE], SRI VANI EDUCATIONAL SOCIETY GROUP
OF INSTITUTIONS, A.P., India.

VADAPALLI GOPI, Associate Professor & HEAD OF THE DEPARTMENT, Dept of CSE,
SRI VANI EDUCATIONAL SOCIETY GROUP OF INSTITUTIONS, A.P., India.

Abstract —However, as payments using a credit card has been increased, frauds associated with it also been increased. The enhancement in technology and availability of information has opened many doors for fraudsters to steal credit card details and do frauds. Due to these frauds, Banks, Companies, Product vendors and Cardholders are largely affected and may face huge financial loss.

INTRODUCTION

Nowadays, Credit cards are widely used for online transactions. Just by having credit card details is more than enough for making online transactions. Thus, the credit card has made online transactions more convenient and easier. Credit card fraud is increasing day by day. Credit card fraud can be done in both online and offline transactions. In offline transactions Physical cards are required while in online transactions the virtual cards are required for doing illegal or fraud activities. Thus these fraud activities in credit card may lead to many fraud transactions without the knowledge of the actual users. The

fraudsters are looking for sensitive information such as credit card number, bank account and other user details in order to perform transactions. In case of offline transactions the fraudsters has to steal the credit card of the user to do the transactions and for the online transactions the fraudsters has to steal the user's identity and online details to perform the online transactions. Thus the credit card fraud has become the major issue in today's technological world which has a massive problem in bank transactions. There are many fraud transactions which cannot be easily identified by the user and also by the banking authority which leads to loss of sensitive data. There are various models which are used for detecting the fraud transactions based on the behavior of the transactions and these methods can be classified as two broad categories such as supervised learning and unsupervised learning algorithm. In existing system for finding the accuracy of the fraudulent activates they have used methods such as Cluster Analysis, Support Vector Machine, Naïve Bayer's Classification etc.

LITERATURE SURVEY

[1] Research on Credit Card Fraud Detection Model Based on Distance Sum. “Wen-Fang YU, Na Wang”.

Along with increasing credit cards and growing trade volume in China, credit card fraud rises sharply. How to enhance the detection and prevention of credit card fraud becomes the focus of risk control of banks. It proposes a credit card fraud detection model using outlier detection based on distance sum according to the infrequency and unconventionality of fraud in credit card transaction data, applying outlier mining into credit card fraud detection. Experiments show that this model is feasible and accurate in detecting credit card fraud.

[2] Fraudulent Detection in Credit Card System Using SVM & Decision Tree. “Vijayshree B. Nipane, Poonam S. Kalinge, DipaliVidhate, Kunal War, Bhagyashree P. Deshpande”.

With growing advancement in the electronic commerce field, fraud is spreading all over the world, causing major financial losses. In current scenario, Major cause of financial losses is credit card fraud; it not only affects trades person but also individual clients. Decision tree, Genetic algorithm, Meta learning strategy, neural network, HMM

are the presented methods used to detect credit card frauds. In contemplate system for fraudulent detection, artificial intelligence concept of Support Vector Machine (SVM) & decision tree is being used to solve the problem. Thus by implementation of this hybrid approach, financial losses can be reduced to greater extend.

[3] Supervised Machine (SVM) Learning for Credit Card Fraud Detection. “Sitaram patel, Sunita Gond”.

This thesis propose the SVM (Support Vector Machine) based method with multiple kernel involvement which also includes several fields of user profile instead of only spending profile. The simulation result shows improvement in TP (true positive), TN (true negative) rate, & also decreases the FP (false positive) & FN (false negative) rate.

[4] Detecting Credit Card Fraud by Decision Trees and Support Vector Machines. “Y. Sahin and E. Duman”

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.

PROPOSED SYSTEM

In existing System, a research about a case study involving credit card fraud detection, where data normalization is applied before Naïve Bayes and Cluster Analysis and with results obtained from the use of these methods on fraud detection has shown that by clustering attributes neuronal inputs can be minimized and promising results can be obtained by using normalized data. 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%. Thus the accuracy of the results obtained from these methods are less when compared with the proposed system.

Demerits of ExistingSystem

1) The algorithm has few downsides such as inefficiency to handle the categorical variables which has different number of levels. Also, when there is an increase in the number of trees, the algorithm's time efficiency takes a hit.

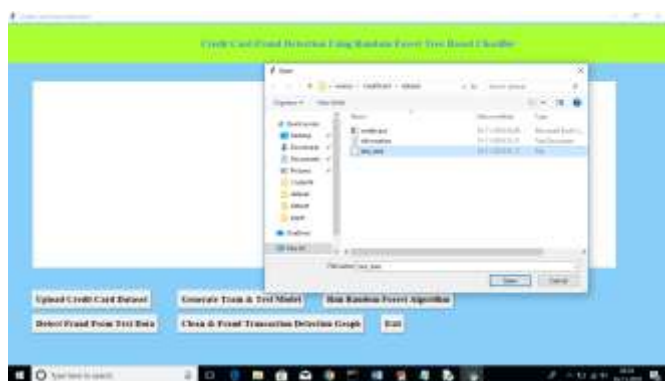
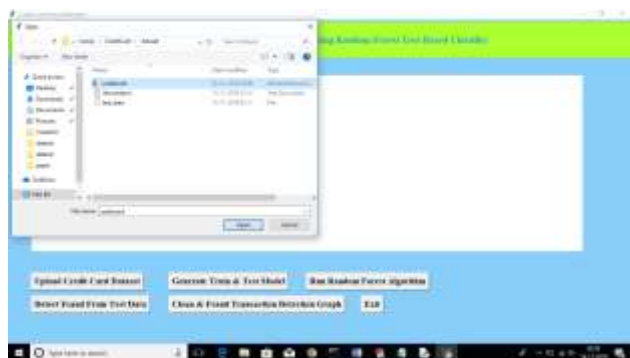
Proposed System

In Proposed system we use Random Forest Algorithm and Neural Networks for classification and regression of dataset. First we will collect the Credit Card dataset and analysis will be done on the collected dataset. After the analysis of dataset then cleaning of dataset is required. Generally in any dataset there will be many duplicate and null values will be present, so to remove all those duplicate and null values cleaning process is required. Then we have to split the dataset into two categories as Trained dataset and Testing dataset for comparing and analyzing the dataset. After dividing the dataset we have to apply the Random Forest Algorithm where this algorithm will gives us the better accuracy about the credit card fraud transactions. By applying the Random Forest Algorithm the dataset will be classified into four categories which will be obtained in the form of confusion matrix. Based on the above classification of data performance analysis will be done. In this analysis the accuracy of credit card fraud transactions can be obtained which will be finally represented in the form of graphical representation.

Merits of Proposed System

The random forest ranks the significance of variables in a natural way that Random Forest can be done in a regression or classification problem. The transaction sum is the 'amount' feature. The 'Class' function is the binary class target class and takes value 1 for positive cases (fraud) and value 0 for negative situations (not fraud).

SAMPLE OUTPUT SCREENSHOTS



CONCLUSION

In this paper, It is helpful to timely find out the students who need to focus on the large amount of campus network usage, to take timely educational assistance measures to help them get a healthy schedule and active learning. Student campus behavior data mining can help teachers and counselors engaged in student affairs management to objectively and timely understand the status and behavior of student groups, and target students with potential problems that need attention. However, the data reveals only "effects". "Cause" also requires face-to-face communication. Student counselor work is essentially a people's work. The exchange of ideas and collisions is an area where data is difficult to describe. Therefore, it is necessary to dialectically apply data mining results to make use of big data technology supports the data of student work, and constantly improves the professional and scientific level of student counselor work.

REFERENCES

- [1] Ramsay J. O., Silverman B. W. Functional data analysis[M]. New York: Springer, 1997.
- [2] Ramsay J. O., Silverman B. W. Applied functional data analysis: methods and case studies. Vol. 77. New York: Springer, 2002.
- [3] KeshengLiu, SiyangWang. Variable selection in regression models including functional data predictors. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 1990-1994.

- [4] Romero C., Ventura S., Data mining in education[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2013, 3(1):12-27.
- [5] Locantore N., Marron J., Simpson D., et al. Robust principal component analysis for functional data[J]. Test, 1999, 8(1):1-73.
- [6] Yao F., Lee T. Penalized spline models for functional principal component analysis[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2006, 68(1):3-25.
- [7] Media.w2018 Chinese College Students' Online Leisure and Entertainment Behavior Monitoring Analysis Reportx[EB/OL]. 2018-11-16. <https://www.iimedia.cn/c400/62969.html>.
- [8]. T.T.Frantz, B.Snider, "Dimensions of Student Behavior", The Journal of Educational Research, vol 63(4), 1969, pp 181-184.
- [9]. K.Bicikova, "Understanding Student Travel Behavior: A Segmentation Analysis of British University Students." , Journal of Travel and Tourism Marketing, Vol 31(7), October 2014, pp. 854-867
- [10]. Shoham, C.Schrage, S. v. Eeden, "Student Travel Behavior", Journal of Travel and Tourism Marketing , Vol 17(4) 2004, pp 1-10.
- [11]. L.Kashunda Williams PhDa, H.George Noell PhDb, A.Beth Jones PhDc and A. Kristin Gansle PhDd, "Modifying Students' Classroom Behaviors Using an Electronic Daily Behavior Report Card " Child and Family Behavior Therapy, Vol 34(4), October 2012