ISSN: 2366-1313

CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING

[1]Sharada, [2]P. Bhavana reddy, [3]R. Maha lakshmi, [4]D. Indu, [5]N. Satwika

[1]Assistant Professor, Department of CSE(DS), Malla Reddy Engineering College for Women (Autonomous Institution – UGC, Govt. of India), Hyderabad, INDIA.

[2345]UG,Department of CSE(DS), Malla Reddy Engineering College for Women (Autonomous Institution – UGC, Govt. of India), Hyderabad , INDIA.

Abstract:

With rapid advancement in the E-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. Credit card frauds are easy and friendly targets. E-commerce and many other online sites have increased the online payment modes, increasing the risk for online frauds. In recent years, For banks has become very difficult for detecting the fraud in credit card system. Machine Learning(ML) plays a important key role for detecting the credit card fraud in the transactions. The main address of the research is to design and develop a fraud detection method for Streaming Transaction Data, with an objective, to analyse the past transaction details of the customers and extract the behavioural patterns. the proposed system is implemented with Support vector machine (SVM) classification to detect the frauds. The conclusion of our study explains the best classifier by training and testing using supervised techniques that provides better solution.

Keywords: Credit Card, Machine Learning, Supervised Technique, Support Vector Machine.

1 INTRODUCTION

In today's world the credit card fraud is the biggest issue and now there is need to fight against the credit card fraud. "credit card fraud is the process of cleaning dirty money, there by making the source of funds no longer identifiable." The purpose may be to obtain goodies without paying, or to obtain unauthorized funds from an account or to avail some kind of services. Credit card fraud is also an add on to identity theft. On daily basis, the financial transactions are made on huge amount in global market and hencedetecting credit card fraud activity is challenging task. The promising way to detectthe fraud is to analyze the spending behavior of the cardholder. Every day, new and new researches are performed by the researchers in the different fields. Many researchers of finance field considered this problem as

ISSN: 2366-1313

a challenging and important problem. The use of machine learning is proposed by the researchers to deal with this problem. Detecting the fraud means identifying the suspicious one, If any abnormality arises in the spending behavior then it is considered as suspicious. This research is to propose a credit card fraud detection system using supervised learning algorithm. supervised algorithms are evolutionary algorithms which aim at obtaining better solutions as time progresses. To Overcomes issues of we propose Machine learning method using 'Structural Similarity', to identify common attributes and behavior with other bank account transaction. Detection of credit card fraud transaction from large volume dataset is difficult, so we propose case reduction methods to reduces the input dataset and then find pair of transaction with other bank account with common attributes and behavior. To elude computational complexity & to provide better accuracy in fraud detection in proposed work. Support vector machine(SVM) is a method used in pattern recognition & classification. It is a classifier to predict or to classify patterns into two categories which may be fraudulent or non fraudulent.

A credit card is a thin handy plastic card that contains identification information such as a signature or picture, and authorizes the person named on it to charge purchases or services to his account - charges for which he will be billed periodically. Today, the information on the card is read by automated teller machines (ATMs), store readers, bank and is also used in online internet banking system. They have a unique card number which is of utmost importance. Its security relies on the physical security of the plastic card as well as the privacy of the credit card number. There is a rapid growth in the number of credit card transactions which has led to a substantial rise in fraudulent activities. Credit card fraud is a wide-ranging term for theft and fraud committed using a credit card as a fraudulent source of funds in a given transaction. Generally, Most of the credit card fraud detection systems are based on artificial intelligence, Meta learning and pattern matching.

2 LITERATURE SURVEY

1] Vimala Devi. J et al. To detect counterfeit transactions, three machine-learning algorithms were presented and implemented. There are many measures used to evaluate the performance of classifiers or predictors, such as the

ISSN: 2366-1313

Vector Machine, Random Forest, and Decision Tree. These metrics are either prevalencedependent or prevalence-independent. Furthermore, these techniques are used in credit card fraud detection mechanisms, and the results of these algorithms have been compared. 2] Popat and Chaudhary. supervised algorithms were presented Deep learning, Logistic Regression, Nave Bayesian, Support Vector Machine (SVM), Neural Network, Artificial Immune System, K Nearest Neighbour, Data Mining, Decision Tree, Fuzzy logic based System, and Genetic Algorithm are some of the techniques used. We compared machine-learning algorithms to prediction, clustering, and outlier detection. 3] Deepa and Akila . For fraud detection, different algorithms like Anomaly Detection Algorithm, K-Nearest Neighbor, Random Forest, K-Means and Decision Tree were used. Based on a given scenario, presented several techniques and predicted the best algorithm to detect deceitful transactions. To predict the fraud result, the system used various rules and algorithms to generate the Fraud score for that certain transaction. 4] Kibria and Sevkli. Using the grid search technique, create a deep learning model. The built model's performance is compared to the performance of two other traditional machine-learning algorithms: logistic regression (LR) and support vector machine (SVM). The developed model is applied to the credit card data set and the results are compared to logistic regression and support vector machine models. 5] Borse Suhas and Dhotre Machine learning's Naive Bayes classification was used to predict common or fraudulent transactions. 6] Asha R B et al. have proposed a deep learning-based method for detecting fraud in credit card transactions. Using machine-learning algorithms such as support vector machine, k-nearest neighbor, and artificial neural network to predict the occurrence of fraud.

3 METHODOLOGY

1) Logistic Regression: Logistic regression works with sigmoid function because the sigmoid function can be used to classify the output that is dependent feature and it uses the probability for classification of the dependent feature. This algorithm works well with less amount of data set because of the use of sigmoid function if value the of sigmoid function is greater than 0.5 the output will 1 if the output the sigmoid function is less than 0.5 then the output is considered as the 0. But this sigmoid function is not suitable for deep learning because the if deep

ISSN: 2366-1313

learning when we back tracking from the output to input we have to update the weights to minimize the error in weight update. we have to do differentiation of sigmoid activation function in middle layer neuron then results in the value of 0.25 this will affect the accuracy of the module in deep learning.

2) Decision Tree: Decision tree can be used for the classification and regression problems working for both is same but some formulas will change. Classification problem uses the entropy and information gain for the building of the decision tree model. entropy tell about how the data is random and information gain tells about how much information we can get from this feature. Regression problem uses the gini and gini index for the building of the decision tree model. In classification problems the root node is selected by using information gain that the root node t id selected by using is having the high information again and low entropy. In Regression problems the root node is selected by using gini , the feature which is having the less gini is selected as the root here Depth of the tree can be determined by using hyper parameter optimization, this can be achieved by Using grid search cv algorithm.

3) Random Forest: The random forest randomly selects the features that is independent variables and also randomly selects the rows by row sampling and the number of decision tree can be determined by using hyper parameter optimization. For classification problem statement the output is the maximum occurrence outputs from each decision tree models inside the random forest. This is one the widely used machine learning algorithm in real word scenarios and in deployed models. And in most of the Kaggle computation challenges this algorithm is used to solve the problem statement.

4 RESULTS

This section presents the results and discussion from our proposed approach and compares the performance of developed hybrid models to the state-of-the-art machine learning algorithms, namely LR, RF, DT, XGBOOST, SVM, NB, Adaboost and LGBM. The single algorithms were compared in terms of prediction performance using their AUROC score to find which ones perform the best in this dataset and therefore are most suited for use as the first algorithm for the proposed hybrid models. The decision was made based on the highest TPR and lowest FPR achieved by Adaboost, while NB gives

ISSN: 2366-1313

the worst performance with an AUROC score of 0.56. The low performance of NB relies completely on the independence assumptions, whereas the used dataset might have some dependence features. However, it demonstrated one of the highest performance rates in the AUC-PR alongside SVM and LGBM. One the other hand, DT and LR has shown the worst performance with 0.22 and 0.28 AUC-PR measure, respectively.

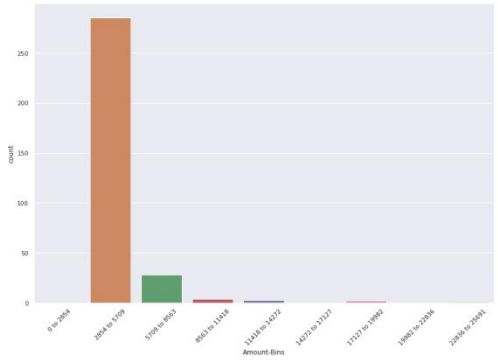

Fig 1: Terms of the card

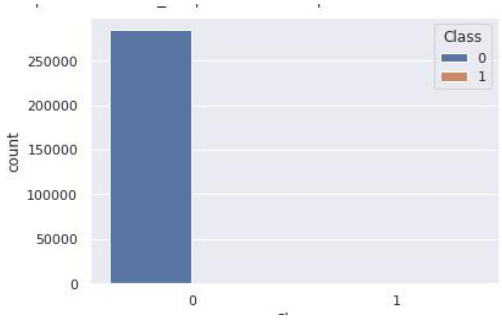

Fig 1: Class labels of the dataset

In the second phase, seven hybrid machine learning models were developed (**Figure 9**). The predictive performance of the hybrid models has shown that the performance of the hybrid model Adaboost + LGBM excels in terms of its AUROC measure when utilizing real-world dataset (IEEE– CIS). The experimental results show that most of our proposed approaches outperformed the state-of-the-art machine learning algorithms in terms of AUROC, Type-I error, Type-II error, F1-measure, precision, misclassification rate and TNR, although some of the hybrid models (Adaboost +LR, Adaboost + NB and Adaboost + SVM) had a higher Type-II error than the state-of-the-art algorithms. However, this will not be a server issue as Type-I error is more costly and being able to lower such an error will have a good impact on the bank system. Additionally, all the proposed hybrid models were able to detect the non-fraudulent cases that were identified as non-fraud at a rate of almost 0.99 percent.

5 Conclusions

Credit card fraud has recently become a major concern worldwide, especially for financial institutions. Various approaches have been previously used to detect fraudulent activities; however, the need to investigate different reliable methods still exists to detect fraudulent credit card transactions, as was the aim in this work for a single case study. In this research, several hybrid machine

ISSN: 2366-1313

learning models were developed and investigated based on the combination of supervised machine learning techniques as a part of a credit card fraud detection study. The hybridization of different models was found to have the ability to yield a major advantage over the state-of-the-art models. However, not all hybrid models worked well with the given dataset. Several experiments need to be conducted to examine various types of models to define which works the best. Comparing the performance of the hybrid model to the state-of-the-art and itself, we conclude that Adaboost + LGBM is the champion model for this dataset. The result also illustrates that the use of hybrid methods has lowered the error rate. For future work, the hybrid models used in this study will be extended to other datasets in the credit card fraud detection domain.

Future work may focus on different areas, starting by proposing data preprocessing techniques to overcome the drawback of the missing values. Additionally, different methods of feature selection and extraction should be investigated in the credit card domain and to determine its impact on prediction accuracy. An investigation of the most appropriate hybrid model among the state-of-the-art machine learning algorithms to determine the most accurate hybridized model in the previously mentioned domain should be the main concern for future studies.

REFERENCES

[1] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and G. N. Surname,"Random forest for credit card fraud detection", IEEE 15th International Conference on Networking, Sensing and Control (ICNSC),2018.

[2] Satvik Vats, Surya Kant Dubey, Naveen Kumar Pandey, "A Tool for Effective Detection of Fraud in Credit Card System", published in International Journal of Communication Network Security ISSN: 2231 – 1882, Volume-2, Issue-1, 2013.

[3] Rinky D. Patel and Dheeraj Kumar Singh, "Credit Card Fraud Detection & Prevention of Fraud Using Genetic Algorithm", published by International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-6, January 2013.

[4] M. Hamdi Ozcelik, Ekrem Duman, Mine Isik, Tugba Cevik, "Improving a credit card fraud detection system using genetic algorithm", published by International conference on Networking and information technology, 2010.

[5] Wen-Fang YU, Na Wang," Research on Credit Card Fraud Detection Model Based on Distance Sum", published by

ISSN: 2366-1313

IEEE International Joint Conference on Artificial Intelligence, 2009.

[6] Andreas L. Prodromidis and Salvatore J. Stolfo; "Agent-Based Distributed Learning Applied to Fraud Detection"; Department of Computer Science- Columbia University; 2000.

[7] Salvatore J. Stolfo, Wei Fan, Wenke Lee and Andreas L. Prodromidis; "Cost-based Modeling for Fraud and Intrusion Detection: Results from the JAM Project"; 0-7695-0490-6/99, 1999 IEEE.

[8] Soltani, N., Akbari, M.K., SargolzaeiJavan, M., "A new user-based model for credit card fraud detection based on artificial immune system," Artificial Intelligence and Signal Processing (AISP), 201216th CSI International Symposium on., IEEE, pp. 029-033, 2012.

[9] S. Ghosh and D. L. Reilly, "Credit card fraud detection with a neuralnetwork", Proceedings of the 27th Annual Conference on System Science, Volume 3: Information Systems: DSS/ Knowledge Based Systems, pages 621-630, 1994. IEEE Computer Society Press.

[10] MasoumehZareapoor, Seeja.K.R, M.Afshar.Alam, "Analysis of Credit Card Fraud Detection Techniques: based on Certain Design Criteria", International Journal of Computer Applications (0975 – 8887) Volume 52– No.3, 2012.

[11] Fraud Brief – AVS and CVM, Clear Commerce Corporation, 2003, http://www.clearcommerce.com.

[12] All points protection: One sure strategy to control fraud, Fair Isaac, http://www.fairisaac.com, 2007. [13] Clear Commerce fraud prevention guide, Clear Commerce Corporation, 2002,http://www.clearcommerce.com

[13] Samaneh Sorournejad, Zahra Zojaji , Reza Ebrahimi Atani , Amir Hassan Monadjemi, "A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective ", IEEE 2016