Journal of Engineering Sciences

Vol 10, Issue 9, Sept /2019
ISSN NO:0377-9254

CREDIT CARD FRAUD DETECTION USING ADABOOST AND MAJORITY VOTING

[#1]B.MONIKA, *M.Tech Student*,

[#2]Dr.G.PRABHAKAR RAO, *Associate Professor*,

DEPARTMENT OF CSE ,

SCIENT INSTITUTE OF TECHNOLOGY,IBRAHIMPATNAM, RANGAREDDY,T.S.

Abstract: The credit card fraud is mostly come in financial services. The credit card fraud is generated huge number of problems in every year. Lack of research on this credit card problem and submits the real-world credit card fraud analyzes, that is issues. In this paper is introduced best data mining algorithm called "machine learning algorithm", which is used to detect the credit card fraud, so initially use this algorithm and it is one of the standard model. Then, hybrid methods which use AdaBoost and majority voting methods are applied. To evaluate the model efficacy, a publicly available credit card data set is used. Then, a real world credit card data set from a financial institution is analyzed. In addition, noise is added to the data samples to further assess the robustness of the algorithms. The experimental results positively indicate that the majority voting method achieves good accuracy rates in detecting fraud cases in credit cards.

Index terms :- credit card, fraud detection, electronic transaction, AdaBoost, majority voting, classification.

I.INTRODUCTION

Credit-card-based purchases can be categorized. In to two types:1)physicalcardand2)virtual card. 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 [1]. Invention of credit cards has made online transactions seamless, comfortable and convenient. Credit card fraud is concerned with the illegal use of credit card information for purchases. Credit card transactions can be accomplished either physically or digitally. In physicaltrans actions, the credit card is involved during the transactions. In digital transactions, this can happen over the telephone or the internet. Cardholders typically provide the card number, expiry date, and card verification number through telephone or website.

With the rise of e-commerce in the past decade, the use of credit cards has increased dramatically. The number of credit card transactions in 2011 in Malaysia were at about 320 million, and increased in 2015 to about 360 million. Along with the rise of credit card usage, the number of fraud cases have been constantly increased. While numerous authorization techniques have been in place, credit card fraud cases have not hindered effectively. Fraudsters favour the internet as their identity and location are hidden. The rise in credit card fraud has a big impact on the financial industry. The global credit card fraud in 2015 reached to a staggering USD $21.84 billion. effective fraud detection system to reduce or eliminate fraud cases is important. There have been various studies on credit card fraud detection. Machine learning and related methods are most commonly used, which include artificial neural networks, rule induction techniques, decision trees, logistic regression, and support vector machines [1]. These methods are used either standalone or by combining several methods together to form hybrid models.

In this paper, a total of twelve machine learning algorithms are used for detecting credit card fraud. The algorithms range from standard neural networks to deep learning models. They are evaluated using both benchmark and real world credit card data sets. In addition, the AdaBoost and majority voting methods are applied for forming hybrid models. To further evaluate the robustness and reliability of the models, noise is added to the real-world data set. The key contribution of this paper is the evaluation of a variety of machine learning models with a real-world credit card data set for fraud detection. While other researchers have used various methods on publicly available data sets, the data set used in this paper are extracted from actual credit card transaction information over three months.

The organization of this paper is as follows. In SectionII, related studies on single and hybrid machine learning algorithms for financial applications is given. The machine learning algorithms used in this study are presented in Section III.The experiments with both benchmark and real world credit card data sets are presented in Section IV. Concluding remarks and recommendations for further work are given in Section V.

II. RELATED WORK

In this section, single and hybrid machine learning algorithms for financial applications are reviewed. Various financial applications from credit card fraud to financial statementfraud are reviewed.

A. SINGLE MODELS

For credit card fraud detection, Random Forest (RF), Support Vector Machine, (SVM) and Logistic Regression (LOR) were examined in. The data set consisted of one-year transactions. Data under-sampling was used to examine the algorithm performances, with RF demonstrating a better performance as compared with SVM and LOR [6]. An Artificial Immune Recognition System (AIRS) for credit card fraud detection was proposed in. AIRS is an improvement over the standard AIS model, where negative selection was used to achieve higher precision. This resulted in an increase of accuracy by 25% and reduced system response time by 40%.

A credit card fraud detection system was proposed in, which consisted of a rule-based filter, Dumpster–Shafer adder, transaction history database, and Bayesian learner. The Dempster–Shafer theory combined various evidential information and created an initial belief, which was used to classify a transaction as normal, suspicious, or abnormal. If a transaction was suspicious, the belief was further evaluated using transaction history from Bayesian learning.

B. HYBRID MODELS

Hybrid models are combination of multiple individual models. A hybrid model consisting of the Multilayer Perceptron (MLP) neural network, SVM, LOR, and Harmony Search (HS) optimization was used in to detect corporate tax evasion. HS was useful for finding the best parameters for the classification models. Using data from the food and textile sectors in Iran, the MLP with HS optimization acquired the highest accuracy rates at 90.07%. A hybrid clustering system with outlier detection capability was used to detect fraud in lottery and online games. The system aggregated online algorithms with statistical information from the input data to identify a number of fraud types. The training data set was compressed into the main memory while new data samples could be incrementally added into the stored data cubes. The system achieved a high detection rate at 98%, with a 0.1% false alarm rate. To tackle financial distress, clustering and classifier ensemble methods were used to form hybrid models in. The SOM and k-means algorithms were used for clustering, while LOR, MLP, and DT were used for classification. Based on these methods, a total of 21 hybrid models with different combinations were created and evaluated with the data set. The SOM with the MLP classifier performed the best, yielding the highest prediction accuracy. An integration of multiple models, i.e. RF, DR, Roush Set Theory (RST), and back-propagation neural network was used in to build a fraud detection model for corporate financial statements. Company financial statements in period of 1998 to 2008 were used as the data set. The results showed that the hybrid model of RF and RST gave the highest classification accuracy.

III. PROPOSED METHOD

To overcome the limitations of existing technology in this paper, a total of twelve machine learning algorithms are used for detecting credit card fraud. The algorithms range from standard neural networks to deep learning models.

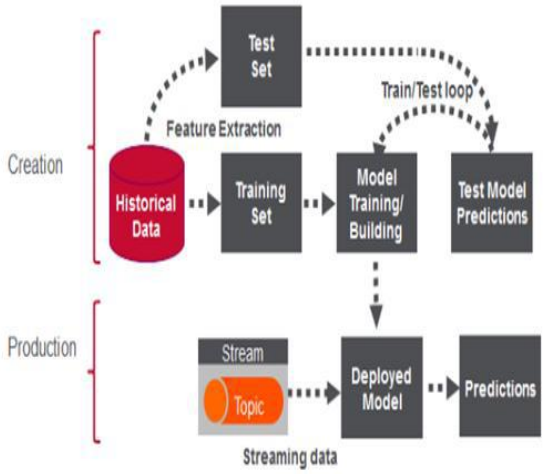

Fig. 1: Architecture of proposed system

They are evaluated using both benchmark and real-world credit card data sets. In addition, the AdaBoost and majority voting methods are applied for forming hybrid models. To further evaluate the robustness and reliability of the models, noise is added to the real-world data set. The key contribution of this paper is the evaluation of a variety of machine learning models with a real-world credit card data set for fraud detection. While other researchers have used various methods on publicly available data sets, the data set used in this paper are extracted from actual credit card transaction information over three months.

IV.ALGORITHM

Fraud detection is done using Adaboost and majority voting methods. Adaptive Boosting or Ada Boost is used in conjunction with different types of algorithms to improve their performance. The outputs are combined by using a weighted sum, which represents the combined output of the boosted classifier. AdaBoost tweaks weak learners in favor of misclassified data samples. It is, however, sensitive to noise and outliers. As long as the classifier performance is not random, AdaBoost is able to improve the individual results from different algorithms. AdaBoost helps improve the fraud detection rates, with a noticeable difference for NB, DT, RT, which produce a perfect accuracy rate. The
most significant improvement is achieved by LIR. Majority voting is frequently used in data classification, which involves a combined model with at least two algorithms. Each algorithm makes its own prediction for every test sample. The final output is for the one that receives the majority of the votes. The majority voting method achieves good accuracy rates in detecting fraud cases in credit cards

V. RESULTS

Fig2. Shows registration card details in credit card

Credit card is the most ordinary way to go in a line of credit. Generally, it is provided by a bank or economic favor. The user can enroll the card details from the above fig2.

Fig3. Lodge a complaint about theft.

Nowadays a person's financial account details can be fetched easily due to which credit card frauds have been increased. Hence forth a user can file a complaint with the bank to block the card or the account by the above fig3.

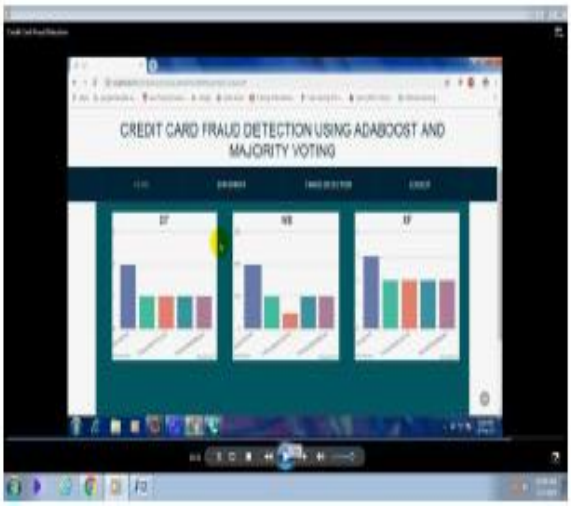

Fig4. Graphical representation of Fraud using Adaboost and Majority Voting

The above graph represents the rate of fraud occurred by online purchase using Majority Voting. And the methods used to detect frauds are: Dessition Tree, Naïve Bayes and Random Forest. X-axis represents the different methods and Y-axis represents the year.

Dark Blue represents the maximum theft occurrence and Red represents the average theft. Whereas Green and Blue represent minimum theft. Finally Purple represents the overall theft occurred throughout the year.

VI. CONCLUSION

A study on credit card fraud detection using machine learning algorithms has been presented in this paper. A number of standard models which include NB, SVM, and DL have been used in the empirical evaluation. A publicly available credit card data set has been used for evaluation using individual (standard) models and hybrid models using Ada Boost and majority voting combination methods. The MCC metric has been adopted as a performance measure, as it takes into account the true and false positive and negative predicted outcomes. A real credit card data set from a financial institution has also been used for evaluation.

For future work, the methods studied in this paper will be extended to online learning models. In addition, other online learning models will be investigated. The use of online learning will enable rapid detection of fraud cases, potentially in real-time. This in turn will help detect and prevent fraudulent transactions before they take place, which will reduce the number of losses incurred every day in the financial sector.

REFERENCES

1. Y. Sahin, S. Bulkan, and E. Duman, "A cost-sensitive decision tree approach for fraud detection," Expert Systems with Applications, vol. 40, no. 15, pp. 5916–5923, 2013.

2. A. O. Adewumi and A. A. Akinyelu, "A survey of machine learning and nature-inspired based credit card fraud detection techniques," International Journal of System Assurance Engineering and Management, vol. 8, pp. 937–953, 2017.

3. A. Srivastava, A. Kundu, S. Sural, A. Majumdar, "Credit card fraud detection using hidden Markov model," IEEE Transactions
on Dependable and Secure Computing, vol. 5, no. 1, pp. 37–48, 2008.

4. The Nilson Report (October 2016) [Online]. Available:https://www.nilsonreport.com/upload/content_promo/The_Nilson_Report_10-17-2016.pdf

5. J. T. Quah, and M. Sriganesh, "Real-time credit card fraud detection using computational intelligence," Expert Systems with Applications, vol. 35, no. 4, pp. 1721–1732, 2008.

6. S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C., "Data mining for credit card fraud: A comparative study," Decision Support Systems, vol. 50, no. 3, pp. 602–613, 2011.

7. N. S. Halvaiee and M. K. Akbari, "A novel model for credit card fraud detection using Artificial Immune Systems," Applied Soft Computing, vol. 24, pp. 40–49, 2014.

8. S. Panigrahi, A. Kundu, S. Sural, and A. K. Majumdar, "Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning," Information Fusion, vol. 10, no. 4, pp. 354–363, 2009.

9. N. Mahmoudi and E. Duman, "Detecting credit card fraud by modified Fisher discriminant analysis," Expert Systems with Applications, vol. 42, no. 5, pp. 2510–2516, 2015.

10. D. Sánchez, M. A. Vila, L. Cerda, and J. M. Serrano, "Association rules applied to credit card fraud detection," Expert Systems with Applications, vol. 36, no. 2, pp. 3630–3640, 2009

11. E. Duman and M. H. Ozcelik, "Detecting credit card fraud by genetic algorithm and scatter search," Expert Systems with Applications, vol. 38, no. 10, pp. 13057–13063, 2011

12. P. Ravisankar, V. Ravi, G. R. Rao, and I. Bose, "Detection of financial statement fraud and feature selection using data mining techniques," Decision Support Systems, vol. 50, no. 2, pp. 491–500, 2011
13. E. Kirkos, C. Spathis, and Y. Manolopoulos, "Data mining techniques for the detection of fraudulent financial statements, "Expert Systems with Applications, vol. 32, no. 4, pp. 995–1003,2007.

14. F. H. Glancy and S. B. Yadav, "A computational model for financial reporting fraud detection," Decision Support Systems,vol. 50, no. 3, pp. 595–601, 2011

15. D. Olszewski, "Fraud detection using self-organizing map visualizing the user profiles," Knowledge-Based Systems, vol. 70, pp. 324–334, 2014

16. J. T. Quah and M. Sriganesh, "Real-time credit card fraud detection using computational intelligence," Expert Systems with Applications, vol. 35, no. 4, pp. 1721–1732, 2008

17. E. Rahimikia, S. Mohammadi, T. Rahmani, and M. Ghazanfari,"Detecting corporate tax evasion using a hybrid intelligent system: A case study of

Iran," International Journal of Accounting Information Systems, vol. 25, pp. 1–17, 2017

18. I. T. Christou, M. Bakopoulos, T. Dimitriou, E. Amolochitis, S.Tsekeridou, and C. Dimitriadis, "Detecting fraud in online games of chance and lotteries," Expert Systems with Applications, vol.38, no. 10, pp. 13158–13169, 2011.

19. C. F. Tsai, "Combining cluster analysis with classifier ensembles to predict financial distress" Information Fusion, vol. 16, pp. 46–58, 2014

20. F. H. Chen, D. J. Chi, and J. Y. Zhu, "Application of Random Forest, Rough Set Theory, Decision Tree and Neural Network to Detect Financial Statement Fraud–Taking Corporate Governance into Consideration," In International Conference on Intelligent Computing, pp. 221–234, Springer, 2014

21.Y. Li, C. Yan, W. Liu, and M. Li, "A principle component
Analysis based random forest with the potential nearest neighbor method for automobile insurance fraud identification," Applied Soft Computing, to be published. DOI: 10.1016/j.asoc.2017.07.027.

AUTHOR'S PROFILE:

[1]. **B.MONIKA,** Pursuing *M.Tech in CSE* **at** Scient Institute Of Technology, Ibrahimpatnam, Rangareddy, T.S.

[2].**Dr. G.PRABHAKAR RAO**, presently working as *Associate Professor in* Department Of CSE, Scient Institute Of Technology, Ibrahimpatnam, Rangareddy,T.S.