International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

SEMISUPERVISED ALGORITHMS BASED CREDITCARD FRAUD DETECTION USING MAJORITY VOTING

[1]G.KUMARA SHEKAR, [2]R.SHIVA RAMAKRISHA

[1]M.TECH DEPT OF CSE, KAKINADA INSTITUTE OF TECHNOLOGICAL SCIENCES, RAMACHANDRAPURAM, ANDHRAPRADESH, INDIA, 533255

[2]ASSISTANT PROFESSOR, KAKINADA INSTITUTE OF TECHNOLOGICAL SCIENCES, RAMACHANDRAPURAM, ANDHRAPRADESH, INDIA, 533255

ABSTRACT

Credit card fraud is a serious problem in financial services. Billions of dollars are lost due to credit card fraud every year. There is a lack of research studies on analyzing real-world credit card data owing to confidentiality issues. In this paper, machine learning algorithms are used to detect credit card fraud. Standard models are firstly used. 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.

1. INTRODUCTION

Fraud is a wrongful or criminal deception aimed to bring financial or personal gain [1]. In avoiding loss from fraud, two mechanisms can be used: fraud prevention and fraud detection. Fraud prevention is a proactive method, where it stops fraud from happening in the first place. On the other hand, fraud detection is needed when a fraudulent transaction is attempted by a fraudster. 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 [2]. In physical transactions, 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 [3]. 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

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

financial industry. The global credit card fraud in 2015 reached to a staggering USD $21.84 billion [4]. Loss from credit card fraud affects the merchants, where they bear all costs, including card issuer fees, charges, and administrative charges [5]. Since the merchants need to bear the loss, some goods are priced higher, or discounts and incentives are reduced. Therefore, it is imperative to reduce the loss, and an 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. IEEE 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 realworld 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.

II. EXISTING SYSTEM

A credit card fraud detection system was proposed in [8], 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 [8]. Simulation results indicated a 98% true positive rate [8]. A modified Fisher Discriminant function was used for credit card fraud detection in [9]. The modification made the traditional functions to become more sensitive to important instances. A weighted average was utilized to calculate variances, which allowed learning of profitable transactions. The results from the modified function confirm it can eventuate more profit [9].

Association rules are utilized for extracting behavior patterns for credit card fraud cases in [10]. The data set focused on retail companies in Chile. Data samples were defuzzified and processed using the Fuzzy Query 2+ data mining tool [10]. The resulting output reduced excessive number of rules, which simplified the task of fraud analysts [10]. To improve the detection of credit card fraud cases, a solution was proposed in [11]. A data set from a Turkish bank was used.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Each transaction was rated as fraudulent or otherwise. The misclassification rates were reduced by using the Genetic Algorithm (GA) and scatter search. The proposed method doubled the performance, as compared with previous results [11].

Disadvantages

There is no Majority Voting technique for credit card fraud detection. There is no Machine Learning Techniques in the existing system.

III. PROPOSED SYSTEM

In the proposed system, 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 is extracted from actual credit card transaction information over three months.

Advantages

The system is very fast due to AdaBoost Technique. Effective Majority Voting techniques.

IV. MODULES

4.1 Bank Admin

In this module, the Admin has to login by using valid user name and password. After login successful he can do some operations such as Bank Admin's Profile ,View Users and Authorize ,View Ecommerce Website Users and Authorize, Add Bank ,View Bank Details ,View Credit Card Requests, View all Products with rank ,View all Financial Frauds ,View all Financial Frauds with Random Forest Tree With wrong CVV ,View all Financial Frauds with Random Forest Tree with Expired Date Usage ,List Of all Users with Majority of Financial Fraud ,Show Product Rank In Chart ,Show Majority Voting With Wrong CVV Fraud in chart ,Show Majority Voting with Expiry date Usage in chart.

4.1.1 View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

4.1.2 View Chart Results

Show Product Rank In Chart, Show Majority Voting With Wrong CVV Fraud in

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

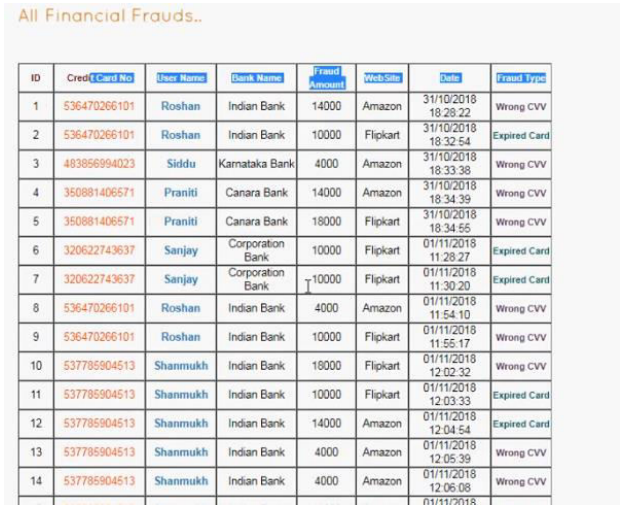
chart, Show Majority Voting with Expiry date Usage in chart.

4.1.3 Ecommerce User

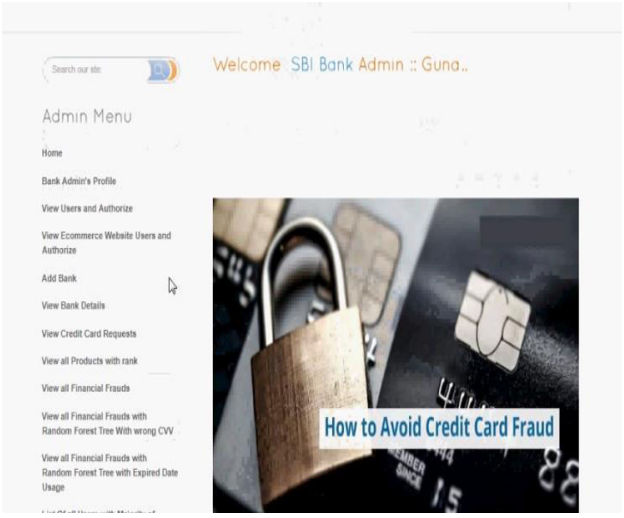
In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like, Add Category, Add Products, View all Products with rank, and View all Purchased Products with total bill, View All Financial Frauds.

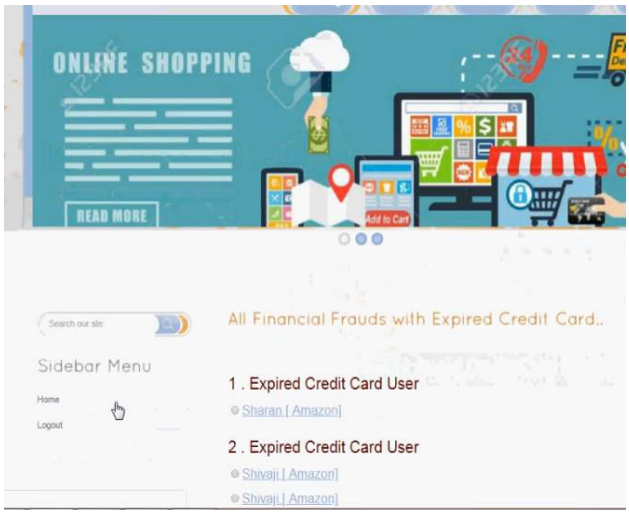
4.2 End User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database.

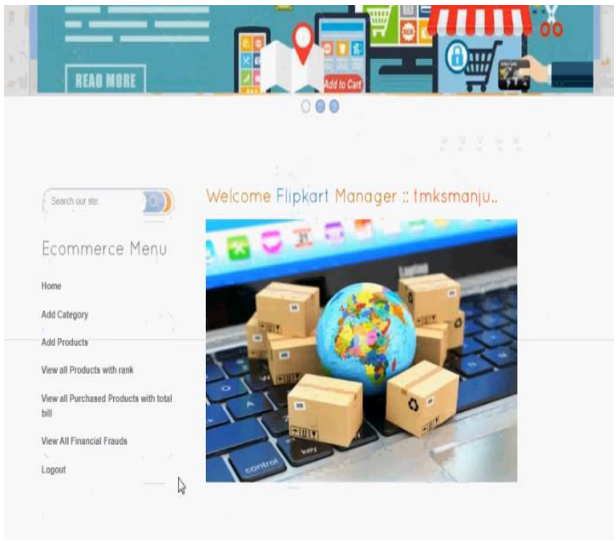


V. SCREEN SHOTOS





International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362



VI. CONCLUSIONS

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 AdaBoost 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. The best MCC score is 0.823, achieved using majority voting. A real credit card data set from a financial institution has also been used for evaluation. The same individual and hybrid models have been employed. A perfect MCC score of 1 has been achieved using AdaBoost and majority voting methods. To further evaluate the hybrid models, noise from 10% to 30% has been added into the data samples. The majority voting method has yielded the best MCC score of 0.942 for 30% noise added to the data set. This shows that the majority voting method is stable in performance in the presence of noise. 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_R eport_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

60techniques," 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 analysisbased 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.

[22] S. Subudhi and S. Panigrahi, "Use of optimized Fuzzy C-Means clustering and supervised classifiers for automobile insurance fraud detection," Journal of King Saud University-Computer and Information Sciences, to be published. DOI: 10.1016/j.jksuci.2017.09.010.

[23] M. Seera, C. P. Lim, K. S. Tan, and W. S. Liew, "Classification of transcranial Doppler signals using individual and ensemble recurrent neural networks," Neurocomputing, vol. 249, pp. 337-344, 2017.

[24] E. Duman, A. Buyukkaya, and I. Elikucuk, "A novel and successful credit card fraud detection system Implemented in a Turkish Bank," In IEEE 13th International Conference on Data Mining Workshops (ICDMW), pp. 162–171, 2013.

[25] C. Phua, K. Smith-Miles, V. Lee, and R. Gayler, "Resilient identity crime detection," IEEE Transactions on Knowledge and Data Engineering, vol. 24, no. 3, pp. 533–546, 2012.

[26] M. W. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation" Journal of Machine Learning Technologies, vol. 2, no. 1, pp. 37–63, 2011.

[27] Credit Card Fraud Detection [Online]. Available:

https://www.kaggle.com/dalpozz/creditcardfraud

[28] R. Saia and S. Carta, "Evaluating Credit Card Transactions in the Frequency Domain for a Proactive Fraud Detection Approach," In Proceedings of the 14th International Joint Conference on eBusiness and Telecommunications, vol. 4, pp. 335–342, 2017.