e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:12/December-2022 Impact Factor- 6.752 www.irjmets.com

MACHINE LEARNING BASED MODELS FOR CREDIT CARD FRAUD DETECTION

Bhargav M[*1], Gaurav G[*2], Harsith S[*3], Suhail Ahmed Sayyed[*4], Mrs. Belji T[*5]

[*1,2,3,4]Student, Department Of Computer Science And Engineering, K. S. School Of Engineering And Management,Bengaluru, Karnataka, India.

[*5]Asst. Prof., Department Of Computer Science And Engineering, K. S. School OfEngineering And Management, Bengaluru, Karnataka, India.

DOI : https://www.doi.org/10.56726/IRJMETS32270

ABSTRACT

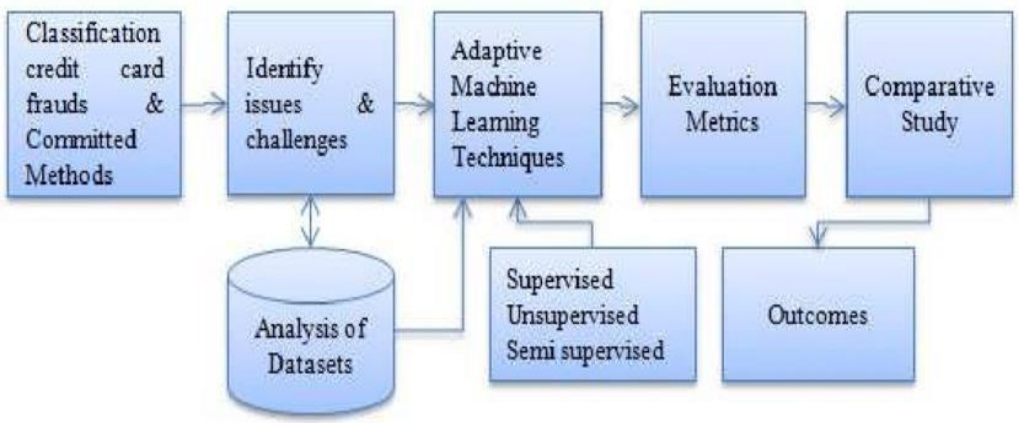
Every year fraud cost generated in the economy is more than $4 trillion internationally. Financial institutions such as commercial and investment banking operations are increasingly bring targeted. Users can use credit card as it provides an efficient and it is easy to use. Due to the increase of usage of credit cards ,the credit card misuse has been enhanced. Fraud detection means a collection of activities to avoid collecting money by misleading pretensions. The main aim to detect such frauds, including the accessibility of public data, the changes in the fraud nature and high rates of false alarm. A machine learning algorithm was first applied to dataset, which improve the accuracy of the detection of frauds to some extent. A number of sectors are today using fraud detection which includes ecommerce and banking agencies. The mode of payment has moved from cash to digital settlements such as debit/credit card, online wallet payment, online banking etc. As the result financial fraud is increasing at rapid rate for personal gain. The algorithms used are K Nearest Neighbors, Random Forest, Decision Tree and Logistic Regression.

Keywords: KNN, Random Forest, Decision Tree, Random Forest, Logistic Regression.

I. INTRODUCTION

Fraud is defined as a wrongful or criminal deception fraud and losses due to fraud. stops fraud from which is aimedbring financial Two mechanisms are used to avoid being happening. fraudulent transaction is attempted by the fraudster. Users can use credit card as it provides an efficientand it is easy to use. aimed bring finance. Users can use creditcard as it provides an efficient and it is easy to use. Due to the increase of usage of credit cards ,the credit card misuse has been enhanced. Fraud detection means a collection of activities to avoid collecting money by misleading pretensions. The main aim to detect such frauds, including the accessibility of public data, the changes in the fraud nature and high rates of false alarm. A machine learning algorithm was first applied to dataset, which improve the accuracy of the detection of frauds to some extent. A number of sectors are today using fraud detection which includes ecommerce and banking agencies. The mode of payment has moved fromcash to digital settlements such as debit/credit card, online wallet payment, online banking etc. As the result financial fraud is increasing at rapid rate for personal gain. The algorithms used are KNN, Random Forest, Decision Tree andLogistic Regression.

II. METHODOLOGY

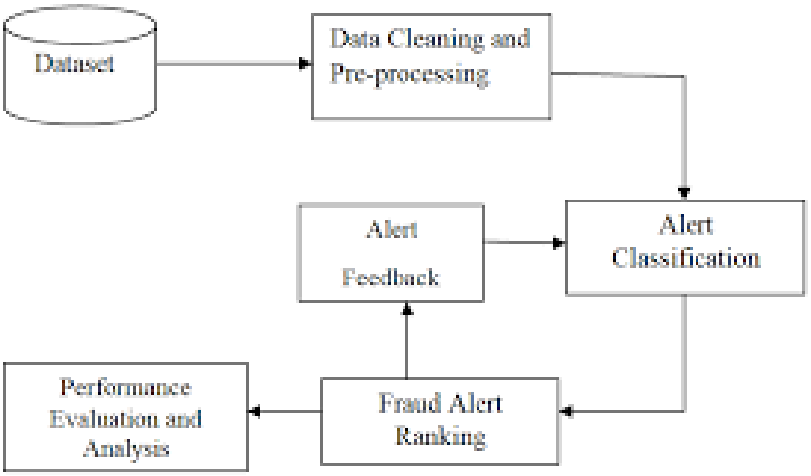


e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:12/December-2022 Impact Factor- 6.752 www.irjmets.com

The flowchart above shows the flow of our application. The different classification techniques we have applied in this study for fraud detection purposes are logistic regression, decision tree, random forest & KNN. Their performances are compared to see which model can better extract the relationship between the features and detect fraudulent transactions. After training all the classifiers, a new ensemble model will be applied as a voting classifier to combine all the other classification techniques. The objective is to reduce the errors of single models, which helps the ensemble model make better predictions compared with the individual classifiers. If all the classifiers are considered as $C1$, $C2$,$C3$, $C4$, and $C5$, then the final classifier will take the votes as the majority of votes as the final prediction or Ct . $Ct = Majority\{C1, C2, C3, C4, C5\}$In the next step.

III. MODELING AND ANALYSIS

The figure below shows the block diagram of our application. In this credit card fraud detection system project. We analyze the given dataset and apply various classification techniques like KNN, Random Forest Model to classify the values as fraudulent or not. We plot a graph and correlation matrix based on this data. We also compare the various machine learning algorithms to find out which is the best classification technique.



IV. RESULTS AND DISCUSSION

In this part, we show the classified result from two prediction models. We used different parameters for make comparison with different models; the parameters i.e. Accuracy, Precision and Recall.

Precision: It is the estimation analysis of true positive to the aggregate value of true positive and false positive rate.

$$\text{Precision} = (TP)/(TP+FP)$$

Recall: It is the estimation analysis of true positive rate to the aggregate value of the true positive and false negative rate.

$$\text{Recall} = (TP)/(TP+FN)$$

The experiments were carried out in two folds. In the first step, a classification process was conducted using F={v1,v2,v3,v4,v5}. For each feature vector in F, the following methods were trained and tested: RF, DT, ANN, NBand LR. The results are depicted. As an initial validation of the proposed method, we ran furtherexperiments using the full feature vector and a feature vector that was generated using a random approach random_vec = { V2, V3, V4, V5, V6, V7, V8, V9, V11, V12, V13, V16, V17, V18, V19, V20, V21, V22, V23, V25, V26, V28, Amount}. Furthermore, we computed the AUC of each vector in F. These results are depicted. The best performing models in terms of the quality of classification are the RF, NB, and LR with the AUCs of 0.96, 0.97, and 0.97, respectively. In the instance of v5 , the RF and NB obtained the highest AUCs of 0.95 and 0.96. Moreover, a comparison analysis is presented in Table 7. This comparison reveals that the GA feature selection approach presented in this paper as well as most of the proposed ML methods that were implemented outperformed the existing techniques that are proposed.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:12/December-2022 Impact Factor- 6.752 www.irjmets.com

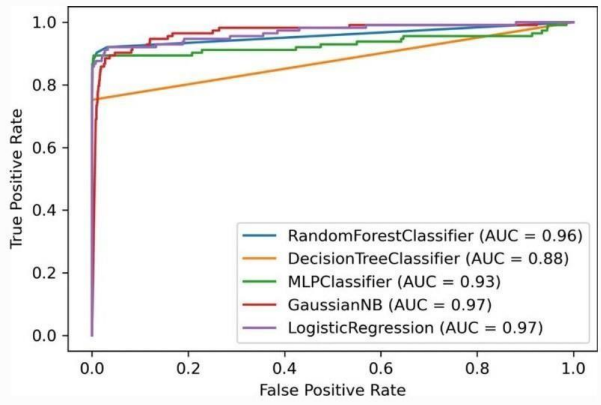


Table 4 Classification results for v_3

From: A machine learning based credit card fraud detection using the GA algorithm for feature selection

Model	Accuracy	Recall	Precision	F1-Score
Rf	99.94 %	75.22 %	85.85 %	80.18 %
DT	99.90 %	76.10 %	68.80 %	72.26 %
ANN	99.91 %	67.25 %	77.55 %	72.03 %
NB	98.81 %	81.41 %	10.07 %	17.93 %
LR	99.90 %	53.09 %	80.00 %	63.82 %

Back to article page >

V. CONCLUSION

Credit card fraud is without a doubt an act of criminal dishonesty. This article has listed out the most common methods of fraud along with their detection methods and reviewed recent findings in this field. This paper has also explained in detail, how machine learning canbe applied to get better results in fraud detection along with the algorithm, pseudocode, explanation its implementation and experimentation results. While the algorithm does reach over 99.6% accuracy, its precision remains only at 28% when a tenth of the data set is taken into consideration. However, when the entire dataset is fed into the algorithm, the precision rises to 33%. This high percentage of accuracy is to be expected due to the huge imbalance between the number of valid and number of genuine transactions. Since the entire dataset consists of only two days' transaction records, its only a fraction of data that can be made availableif this project were to be used on a commercial scale. Being based on machine learning algorithms, the program will onlyincrease its efficiency over time as more data is put into it.

VI. REFERENCES

[1] Research Article Fraud Miner: A Novel Credit Card Fraud Detection Model Based on Frequent Itemset by K.R.Seeja and Masoumeh Zareapoor.

[2] D.Sanchez, M.A.Villa,L. Cerda &J.M.Serrano," Association rules applied to credit card fraud detection", Expert Systems with Applications, vol.36,no.2, pp.3630-3640 .

[3] Baesens, B., Höppner, S., Ortner, I., Verdonck, T., 2021a. robROSE: a robust approach for dealing with imbalanced data in fraud detection. Stat. Methods Appl. doi:10. 1007/s10260-021- 00573.

[4] Baesens, B., Vlasselaer, V.Van., Verbeke, W, 2015. Fraud Analytics Using Descriptive. Predictive. and Social Network Techniques: A Guide to Data Science For Fraud Detection. John Wiley \& Sons. Inc, Hoboken. NJ. USA doi:10.1002/ 9781119146841.

[5] Carrasco, R.S.M., Sicilia-Urbán, M.A., 2020. Evaluation of deep neural networks for reductionof credit card fraud alerts. IEEE Access 8, 186421186432.doi:10.1109/ACCESS.2020.3026222

[6] Kaminski, ´ B., Jakubczyk, M., Szufel, P., 2018. A framework for sensitivity analysis of decision trees. Cent. Eur. J. Oper. Res. 26, 135–159. doi:10.1007/ s10100-017-0479-6.

[7] Sadgali, Imane, Sael, Nawal, Benabbou, Faouzia, 2021. Human behavior scoring in credit card fraud detection. IAES Int. J. Artif. Intell. 10, 698.doi:10.11591/ijai. v10.i3.pp698-706, IJ-AI.

[8] Sun, J., Li, Y., Chen, C., Lee, J., Liu, X., Zhang, Z.,Xu,W., 2020. FDHelper: assist unsupervised fraud detection experts with interactive feature selection and evaluation. In: Paper presented at the Conference on Human Factors in ComputingSystemsProceedingsdoi:10.1145/3313831.3376140.

[9] Trisanto,D.,Rismawati,N.,Mulya,M.,Kurniadi,F., 2020.Effectiveness undersampling method and feature reduction in credit card fraud detection.Int.J.Intell.Eng.Syst.13,173181.
doi:10.22266/ijies2020.0430.17.

[10] Vinod Jain, Mayank Agrawal, Anuj Kumar, "Performance Analysis of Machine Learning Algorithms in Credit Cards Fraud Detection, 2020 at 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2022.

[11] Apapan Pumsirirat, Liu Yan, Credit card fraud detection using deep learning based on auto- encoder and restricted boltzmann machine" in (IJACSA), Int. J. Adv. Comput. Sci. Appl. 9 (1) (201).

[12] Vaishnave Jonnalagadda, Priya Gupta, Eesita Sen in "Credit card fraud detection using Random Forest Algorithm" in Jonnalagadda Vaishnave et al.; International Journal of Advance Research, Ideas and Innovations in Technology, Volume 5, Issue 2.

[13] Amzah Ali Shukur, Sefer Kurnaz in, Credit card fraud detection using machine learning methodology, Int. J. Comput. Sci. Mob. Comput. 8 (3) (2019) 257–260 Vol.Issuepg.

[14] Anuruddha Thennakoon, Chee Bhagyani, Sasitha Premadasa, Shalitha Mihiranga, Nuwan Kuruwitaarachchi in "Real-time Credit Card Fraud Detection Using Machine Learning, 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2022.

[15] Roopesh Akula in "Fraud identification of credit card using ML techniques at, Int. J. Comput. Artif. Intell. 1 (2) (2020) 31–33.