International Journal of Engineering & Technology, 7 (3) (2018) 1083-1086

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i3.12577
Research paper

Fraud prediction for credit card using classification method

Er. Monika[1*], Er. Amarpreet Kaur[2]

[1]*M.Tech Research Scholar, Department of Computer Science and Engineering, Chandigarh University, Punjab, India*

[2]*Assistant Professor, Computer science and Engineering, Chandigarh University, Gharuan, Mohali, Punjab, India*

**Corresponding author E-mail: monikachouhan472@gmail.com*

Abstract

With the improvement of innovation like credit cards, debit cards, mobile banking, Internet managing an account is the mainstream medium to exchange the cash starting with one record then onto the next. Credit card is picking up fame day by day which expands the online exchange with the expansion in online shopping, online charge payment, insurance premium and different charges so the extortion cases identified with this are likewise expanding and it puts an extraordinary anxiety on the economy, affecting the two clients and budgetary bodies. It costs cash as well as an awesome measure of time to reestablish the damage done. In this paper, we look whether data mining procedures are valuable to estimate and categorize the client's credit risk score (normal/fraud) to beat the future dangers. The reason for existing is to keep the clients from online exchange by utilizing particular Data mining classification methods. The fakes are ascertained by Naïve Bayes method way to deal with break down the exchange is actual or fake. The exploratory outcome demonstrates that our model has great classification accuracy, recall and precision.

Keywords: *Classification Approaches; Conventional Neural Network; Data Mining; Fraud detection; Naïve Bayes.*

1. Introduction

Due to advancement in social web-sites, e-commerce, business modelling and various enormous technologies, various huge datasets are available which can be used for predictive modeling [1]. Advancement in communication and information, academics stakeholders are witnessing keen interest in predictive modeling of datasets. Predictive modeling analysis is performed to help academics organizations to drive innovation by acquiring new insight into their customers[2].

Each application that is running today requires large size of databases such that huge amount of data that is being generated within them can be stored in them [3][4]. There is a need to generate new and highly efficient tools which can manage such huge databases and the data present within them. For facilitating the large number of users as well as their applications several techniques have been proposed by researchers [5]. Amongst such techniques is the data mining and knowledge discovery approach that is used in order to manage all such data. Data mining is known as the automated approach of identifying interesting patterns from huge databases[6]. This technique helps in generating descriptive, understandable and predictive models from the existing data within the databases.

The main reason behind the development of this work is to store data which is generated each day due to all the applications has been utilized by humans [7]. In the computer databases, this data is stored so that it becomes easy to access data whenever it is required. The similar objects are separated from each other amongst different groups this process is known as clustering. Cluster is defined as the group in which similar objects are placed within the same group. Similarly, objects having different properties and are dissimilar with each other are placed in separate cluster [8]. The representation of data using less clusters leads to loss of essential information. With the help of this approach, it becomes easy to simply data in controlled manner. In the databases large amount of data is present which is stored from long time. The extraction of this information is very important in order to utilize it in useful manner [9]. The process of data mining extracts the useful information with quality and also provides effective sharing. The discovery of predictive intelligence has been done with the help of uncovering patterns and relationships present within structured as well as unstructured data in case of data mining and text analytics[11-10]. According to researchers, it is required to have deep knowledge or creative skills so that models can be generated within predictive analytics. Hence, this approach is developed with the help of some basic functions. The use of credit card is common nowadays due to change in technology from moderate to modern. Therefore, it is necessary to develop a model for the detection of frauds in credit cards in all regions. Fraud in credit cards is done by any unauthorized account activity of a person[12]. Thus, it is necessary to take some appropriate steps by which increasing frauds and threats can be minimized timely. For the protection of the clients various methods are facilitated that provide the risk management that eliminates the misuses of credit cards misuse. The credit card fraud took place when user as well as issuer has no knowledge that other use tries to use card. Hence, this is known as frauds in which one card is used by another. The person who tries to use card of another person is not having good intention as that person want to withdraw all money from user account [13]. The unauthorized user is identified with the help of fraud detection system, if one could commit the fraud. In order to minimize all these frauds and to attract criminals various fraud detection methods has been developed so far using various strategies. The development of new mechanism for fraud detection is becomes difficult due to the limitations of exchanging the ideas of fraud detection [14]. This study shows how predictive analytics can be implemented using open source technologies and various

Copyright © 2018 Er. Monika and Er. Amarpreet Kaur. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

data mining algorithms to process real time data in fault tolerant manner in scalable and efficient manner.

2. Literature review

K. Randhawa et al. [15] presented the use of machine learning algorithm has been utilized for the detection of the fraud in the credit card by developing a system. The standard models was utilized initially after which the use of hybrid methods is followed in which AdaBoost is present. The efficiency of the model has been evaluated with the help of credit card dataset which is available everywhere. In the data samples noise is embedded in order to evaluate the robustness of algorithms. On the basis of performed experiments, it is concluded that fraud detection methods provides the higher accuracy rates. With the help of this approach online learning models will be proposed in future to extent the work. Hence, it becomes easy to identify fraud cases quickly using online learning mechanism. [16] Presented various issues are faced by individual while selecting appropriate fraud detection models as per done study. The implementation of an effective process coefficient sum mechanism was done whole set of selection criteria and FDMs are introduced. For the analysis of various scenarios, this proposed approach has been utilized using decision maker's criterion. There is no fraud detection model till yet which satisfies all the needs by which frauds can be detected easily. They proposed a method in this paper which is based on the integration of different comparison so that fraud can be detected ranked easily. With the help of this proposed method, the complex multi-attributes decision issues can be resolved easily. [17] Presented the use of fuzzy logic which is the combination of mathematical process by which an ID3 decision tree system is generated in this paper. They implemented the FuzzITree algorithm on the normalized training data. As per performed experiments, it is demonstrated that all the performed transaction was classified correctly except for the transaction number 7. They performed various tests on different transactions so that detection rate can be determined easily in different situations. On the basis of achieved results, it is identified that the detection rate is around 89%. In future, advanced concepts of fuzzy will be utilized so that fraud detection can be identified easily. [18] Presented in order to enhance the performance of credit card fraud detection systems, they proposed various methods which are based on signal processing on graphs. In this paper, they proposed iterative amplitude adjusted Fourier transform along with iterative surrogate signals on graph algorithms as an alternative. It is necessary to do a reliable augmentation of the target inadequate population of frauds due to the present issues. These issues are labeling cost, algorithm testing, data confidentiality and consistent modification of patterns. With the help of legitimate and non-legitimate transaction ratios, they evaluated the detection capabilities feasibility using receiver operating characteristic (ROC) curves and several key performance indicators (KPI) are utilized. [19] Presented, a novel system is purposed for making appropriate decisions related to the approval or rejection of load request of a customer on the basis of various detailed information. This paper analysed performance of various classifiers such as decision tree, support vector machine, adaptive boosting etc. The performance and accuracy achieved through the random forest classification is higher in comparison to rest of the algorithm. [20] Proposed, A RUSMRN algorithm by using machine learning methods. It depends on RUS information inspecting method and MRN calculation for the expectation of the information installment. With a specific end goal to enhance the grouping exactness of unbalance trademark information creator proposed a RUSMRN technique that is a blend of the boosting and information testing. Today's, numerous ventures for the most part centered around the use benefits through charge card extensively on the grounds that it rushes to pay for items and administrations. Creator accentuated on the extortion discovery of Mastercard installment by utilizing the machine learning system called RUSMRN. The proposed technique embraces three base classifiers which are MLP, NB and Naïve Bayes calculations. What's more, it can dissect the rightness to work with the unbalance datasets. Creator presumed that the proposed strategy can accomplish the best grouping execution as far as exactness and affectability. RUSMAN has most elevated affectability in the wake of preparing and testing by applying the propose technique and it is appropriated for foreseeing the information. [21] Presented for the prevention and detection of frauds timely, they proposed a mechanism in this paper. With the help of this method, alert related to fraudulent transactions will be sent to the card owner and credit card is blocked immediately. The value generated for Hidden Markov Model outcome is the basis that decides whether the transaction is fraudulent or not. With the help of this system, all the marked fraud cases among the actual transactions can be minimized using this by comparing the rate of false alarms that are achieved. In order to minimize the false alarm rates and to prevent frauds they utilize the streaming analytics. They proposed a model and update it regularly on the basis of collected information of genuine card holder.

3. Research Methodology

Naive Bayes classifier depends on probabilistic classifier demonstrate that join solid autonomy suspicions which don't have affect on the truth that is the reason these are taken as Naive. This approach utilizes the ordinary circulation to show numeric properties. It likewise dealing with the numeric properties utilizing directed discretization[22]. Naive Bayes is a strategy for creating characterization display that appoints class names to issue occurrences, pointer as vectors of highlight esteems, where the class marks are venture from some limited set. This approach is not for preparing such classifier yet in addition gathering of calculation rely upon normal primary, all guileless Bayes classifier continue that the estimation of any component is autonomous of estimation of some other element for given class marks. Leverage of Naive Bayes classifier is it require a little measure of preparing information to appraise the parameters basic for classification [23][24][25].

The credit card fraud detection algorithm and data classification methods were evaluated by conducting a detailed literature study in the first step. This was done to know their advantages and limitations. The limitation of the existing schemes is overcome by the design of proposed security solution that improves the structure of this method. In order to build a robust system, it combines all the advantages of the system together. For the simulation process, they utilize the PYTHON simulator in which the proposed solution is implemented using all essential input and output parameters. After, implementation of the proposed method they analyze the performance and compared this method with the existing. From last 10 years, the historical data of the credit cards has been utilized for the process of input data acquisition. In the historical data, all the details of transactions of every day is present which provides the information to user when the amount is withdrawn and by what time and from which location.

By collecting the regression models with squared distance, they designed the proposed model during the process of implementation for classification. This was done for processing of historical data which has been utilized for the long term prediction. The classification of the input data into the normal and fraud transactions can be done using Naïve Bayes classifier in this research work. The Naïve Bayes classification is also known as a predictive model which has been utilized for the text categorization. The used input here is data and output is the classified data given in two categories. With the help of Naïve Bayes training algorithm, a model is implemented for the amount of text where each training example belongs to one of the two classes. The data is divided into two categories by constructing N-Dimensional hyperplane. On the each side of the hyper plane, they generated two parallel hyper planes so that data can be separated easily. Hence, by separating the hyper planes, there is increase in the distance between the two hyper planes. In association to the partitioning hyper plane f(X) divides the two classes when it passes through middle. For the creation of the linear classification function, there is a linearly

separable data set. The testing of sign of function f(Xn) easily test and classify the new data instance, Xn when it determine the function.: Where Xn belongs to a positive class if f(Xn) > 0
The generalization of the error of the classifier for the larger distance or margin can be done in an optimal way. This algorithm functioning is very good in concern with high dimensional feature set. For the creation of the new linearly separable data they utilize the kernel trick in which non-linearly separable data is transferred. Naïve Bayes classification has been utilized for the calculation of regression analysis and to perform numerical calculations. This algorithm is also useful in ranking the elements. There are various attributes present on the dataset of the Naïve Bayes and it performance is good as only specific cases can be accessed for training purpose which is considered as the advantage of Naïve Bayes. But, it has limitation of speed and size which occur during the training and testing phase of Naïve Bayes. It is not easy to choose the parameters of kernel function hence, consider as disadvantage of algorithm.

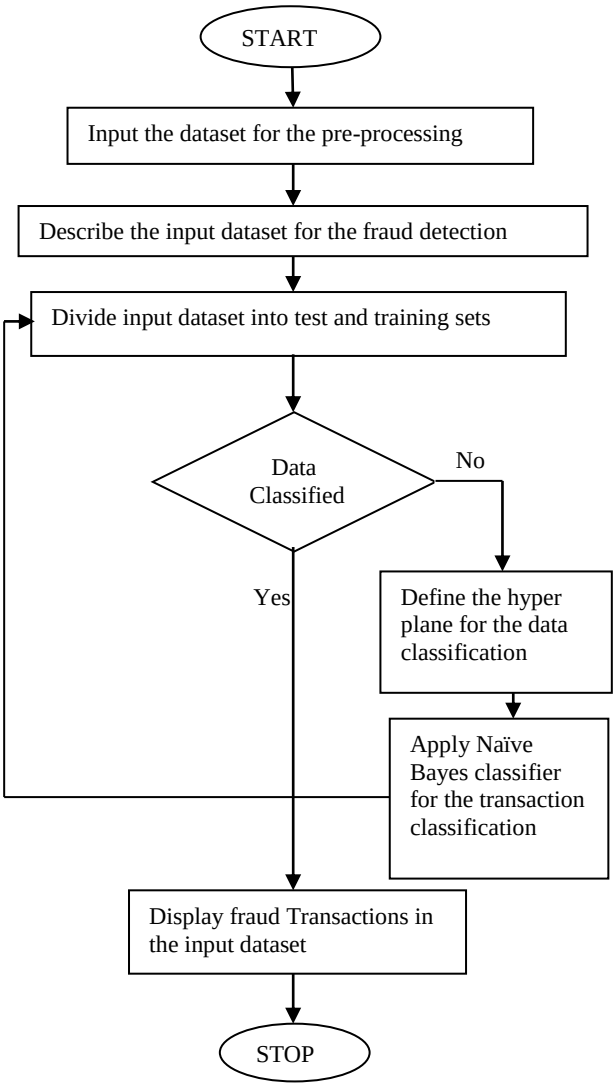


Fig. 1: Proposed Flowchart

4. Results and Discussion

The proposed technique has been implemented in Python and the results have been analyzed in terms of accuracy as shown below. The values of prediction and recall is used to generate the final accuracy values. The values of precision, recall and accuracy is defined in table 1.

Table1: Comparision between different parameters

Parameter	Existing Algorithm	Proposed Algorithm
Precision	89	94
Recall	91	96
Accuracy	84	90

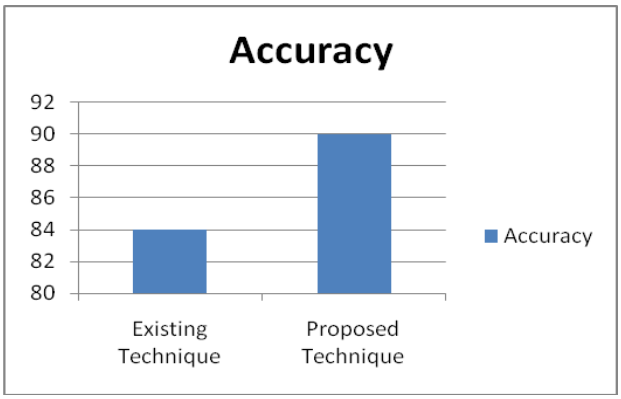


Fig. 2: Accuracy Comparison

As shown in figure 2, the accuracy of proposed and existing algorithms is compared and it is analyzed that proposed algoirhtm performs well in terms of accuracy.

5. Conclusion and Future Scope

Data mining is the technique using which essential information can be easily extracted from the large amount of data. Prediction analysis is considered as the application of credit card fraud detection. The prediction of future data in the prediction analysis has been done using current information. The implementation of neural networks was done in the existing techniques by which input data is learned that drive the future values. For the classification of normal and fraud transactions, they implemented the classification technique in this work. The whole data is divided into test and training sets, by applying the Naïve Bayes classifier in this work. Here, input is the test and training set which provides the future values. Experimental results, evaluated the performance of proposed modal as compared to existing technique in terms of accuracy. The most useful advantage of data mining classification approach is prediction analysis. In this research work, it has been concluded that prediction analysis is the technique which can classify the data according to the set of rules. Predictive analytics can be implemented using open source technologies and various data mining algorithms to process real time data in fault tolerant manner in scalable and efficient manner. Different kind of classification algorithms are used for prediction analysis. However, various classification approaches are used for prediction analysis to improve the accuracy of the predicted model but all the analysis based on a particular dataset which is used to produce a predicted results in terms of Accuracy, Precision, Recall, F-measure, G-measure and so on to find out the bestest classification approach for prediction analysis.

References

[1] N. College, "Performance Analysis of Bayes Classification Algorithms in WEKA Tool using Bank Marketing Dataset," vol. 5, no. 2, pp. 128–133, 2018.
[2] K. C. Tan, E. J. Teoh, Q. Yu, and K. C. Goh, "Expert Systems with Applications A hybrid evolutionary algorithm for attribute selection in data mining," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 8616–8630, 2009.

[3] M. Panda and A. Abraham, “Hybrid evolutionary algorithms for classification data mining,” *Neural Comput. Appl.*, vol. 26, no. 3, pp. 507–523, 2015.

[4] R. Forsati, M. R. Meybodi, M. Mahdavi, and A. G. Neiat, “Hybridization of K-means and harmony search methods for web page clustering,” *Proc. - 2008 IEEE/WIC/ACM Int. Conf. Web Intell. WI 2008*, pp. 329–335, 2008.

[5] Y. Yu, F. Zhong-liang, Z. Xiang-hui, and C. Wen-fang, “Combining Classifier Based on Decision Tree,” *2009 WASE Int. Conf. Inf. Eng.*, pp. 37–40, 2009.

[6] P. Lakhmi Prasanna, D. Rajeswara Rao, Y. Meghana *, K. Maithri, T. Dhinesh, "Analysis of supervised classification techniques", International Journal of Engineering & Technology, 7 (1.1) (2018) 283-285

[7] Salah H.R. Ali, Marwah M.A. Almaatoq and Abdalla S.A. Mohamed, "Classifications, surface characterization and standardization of nanobiomaterials", International Journal of Engineering and Technology, 2 (3) (2013) 187-199

[8] M V.R.Viswanadh, M RameshKumar, T Chandana, "Verification of Certificates Using Smart Card Technology", International Journal of Engineering & Technology, 7 (2.7) (2018) 993-996

[9] Krishnamoorthy. P, Dr. R. Gobinath, "Survey on classifier algorithms for health care system in diabetes", International Journal of Engineering & Technology, 7 (2.26) (2018) 19-24

[10] SriDeivannai Nagarajan, R.M.Chandrasekaran,” Diagnosing Diabetes using data mining Techniques”, International Journal of Engineering Sciences & Research Technology, Volume -5, Pag no 673 679, November- 2015.

[11] Dhivya Selvaraj, Mrs.Merlin Mercy ” Distributed association rule mining and summarization for Diabetes Mellitus and Its Co-Morbid Risk Prediction strategy using FUZZY Classifier”, International Journal of Engineering and Applied Sciences (IJEAS),Volume-2, November 2015.

[12] Ramkumar,Dr.K.Satheshkumar and G.Emayavaramban” Nine States HCI using Electrooculogram and Neural Networks”, IJET, Vol. 8(6), pp. 3056-3064, Jan 2017.

[13] M. Murugesan, R. Elankeerthana, "Support vector machine the most fruitful algorithm for prognosticating heart disorder", International Journal of Engineering & Technology, 7 (2.26) (2018) 48-52

[14] Dr. E. Laxmi Lydia, B. Prasanna Kumar , D. Ramya, "Generation of dynamic energy management using data mining techniques basing on big data analytics isssues in smart grids", International Journal of Engineering & Technology, 7 (2.26) (2018) 85-89

[15] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, “Credit Card Fraud Detection Using AdaBoost and Majority Voting,” *IEEE Access*, vol. 6, pp. 14277–14284, 2018.

[16] S. Arora, “Selection of Optimal Credit Card Fraud Detection Models Using a Coefficient Sum Approach,” pp. 482–487, 2017.

[17] S. S. Askari, “Credit Card Fraud Detection Using Fuzzy ID3,” pp. 446–452, 2017.

[18] L. Vergara, A. Salazar, J. Belda, G. Safont, S. Moral, and S. Iglesias, “Signal processing on graphs for improving automatic credit card fraud detection,” *Proc. - Int. Carnahan Conf. Secur. Technol.*, vol. 2017–October, pp. 1–6, 2017.

[19] A. Gahlaut and P. K. Singh, “Prediction analysis of risky credit using Data mining classification models,” 2017.

[20] A. Charleonnan, “Credit card fraud detection using RUS and MRN algorithms,” *2016 Manag. Innov. Technol. Int. Conf.*, p. MIT-73-MIT-76, 2016.

[21] Rajeshwari U and B. S. Babu, “Real-time credit card fraud detection using Streaming Analytics,” *2016 2nd Int. Conf. Appl. Theor. Comput. Commun. Technol.*, pp. 439–444, 2016.

[22] S. D. Jadhav and H. P. Channe, “Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques,” *Int. J. Sci. Res.*, vol. 14611, no. 1, pp. 2319–7064, 2013.

[23] P. Kaur, M. Singh, and G. S. Josan, “Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector,” *Procedia Comput. Sci.*, vol. 57, pp. 500–508, 2015.

[24] T. J. Peter and K. Somasundaram, “An empirical study on prediction of heart disease using classification data mining techniques,” *IEEE Int. Conf. Adv. Egineering, Sci. Manag.*, pp. 514–518, 2012.

[25] K. Srinivas, G. R. Rao, and A. Govardhan, “Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques,” *2010 5th Int. Conf. Comput. Sci. Educ.*, pp. 1344–1349, 2010.