



CLASSIFICATION TECHNIQUE FOR THE CREDIT CARD FRAUD DETECTION

Tamanna Chouhan¹, Ravi Kant Sahu²

Abstract- The data mining is the technique which can mine useful information from the rough data. The prediction analysis is the technique of data mining which can predict new things from the current data. The classifications techniques are generally applied for the prediction analysis. This research work, is based on the prediction of the credit card fraud detection. The various techniques are proposed by the authors for the credit card fraud detection. The technique which is proposed in the base paper is based on the conventional neural networks in which system learns from the previous experiences and drive new values. In this research work, the SVM classifier is proposed for the credit card fraud detection which can classify input data into normal and fraud transactions. In the proposed technique, input data is divided into test and training sets. On the basis of test and training sets the normal and fraud transactions are predicted in terms of precision and recall.

Keywords – SVC, IAAFT, ROC, ISR

1. INTRODUCTION

There has been an increase in demand of huge databases within every application today due to the growth of data being generated by each second. In order to handle such large databases and manage the data, there is a need to introduce new powerful tools which are highly efficient. Several researchers have proposed numerous techniques in order to facilitate such large number of users and their applications. Thus, a new research area has been developed here which is known as data mining and knowledge discovery. An automated process in which the interesting information or patterns are discovered from large databases is known as data mining [1]. There are descriptive, understandable as well as predictive models generated from the data present within the databases through this approach. In all the applications being used by humans today, there is generation as well as storage of data each day due to which this research work has been developed. In order to provide an ease to access the data as per the requirement, the data is stored within the computer databases. The separation of objects that are similar to each other amongst various groups is known as clustering. The objects that have similarity amongst each other are placed within one group which is also known as cluster [2]. The objects that have different properties are further placed within separate clusters. However, there is loss of specific details when the data is represented with the help of fewer clusters. The advantage here is that the data is simplified in organized manner through this approach. There is huge amount of data present within the information databases which is generated and stored over large time. This information is of complete waste until a user extracts it for some useful purposes [3].

Data mining is a success as per the availability of high quality data as well as effective sharing. In data mining as well as text analytics, with the help of uncovering patterns as well as relationships present within structured as well as unstructured data, the predictive intelligence is discovered [4]. It is stated by many of the researchers that there is a need of great skill as well as an artistic and highly creative side in order to generate models within predictive analytics. However, there are some basic steps also involved which help in developing this approach. Within today's modern society, the utilization of credit cards is very common. Thus, within academic as well as business community, there is a need to develop credit card fraud detection model. An unauthorized account activity of a person in which the account is not intended is known as credit card fraud [5].

Thus, in order to stop the abuse that has been going on and provide risk management facilities within organizations such that the clients can be protected in better way, actions need to be taken against this approach. Thus, a scenario in which one user attempts to use another user's credit card in such a way that the card owner as well as the issuer has no knowledge about it is known as credit card fraud. The intentions of the person who has attempted to use the card are not good and he tries to steal the money of other clients. Once, the fraud has been committed, the fraud detection system makes it possible to identify the unauthorized user as soon as possible. In order to defend criminals by using various strategies, the fraud detection methods have been developed [6]. Due to the severe limitation of exchanging the ideas of fraud detection, it has become more difficult to develop new fraud detection mechanisms. Due to the severe limitation of exchanging ideas within the fraud detection, it has become more difficult to develop new fraud detection methods as well.

¹ Department of Computer Science and Engineering, Lovely Professional University, Jalandhar (Punjab), India

² Department of Computer Science and Engineering, Lovely Professional University, Jalandhar (Punjab), India

2. LITERATURE REVIEW

Kuldeep Randhawa, et.al (2017) proposed a system to detect credit card frauds through the utilization of machine learning algorithms

[7]. Initially, the standard models are utilized which are then followed by hybrid methods that include AdaBoost. A credit card dataset that is publically available is utilized here in order to evaluate the efficacy of model. In order to assess the robustness of algorithms, noise is added to the data samples. Higher accuracy rates are achieved for detection of fraud cases within the credit cards as per the experimental results. The future work will be extended to proposing online learning models through this approach. Within real-time scenarios, the fraud cases can be identified rapidly through the utilization of online learning.

Suman Arora, et.al (2017) presented a study that is related to discussing of issues related to optimal selection of fraud detection models [8]. An effective process coefficient sum mechanism is applied in case when whole set of selection criteria and FDMs are introduced. On the basis of decision maker's preferences, this approach is utilized in order to analyze various scenarios. There is no optimal fraud detection model that satisfies all the selection criteria. On the basis integration of different comparison criteria, a method is proposed in this paper that is appropriate for ranking the fraud detection models. The complex multi-attributes decision issues that involve quantitative as well as qualitative factors are solved through this proposed approach.

S Md. S Askari, et.al (2017) proposed generation of an ID3 decision tree system by utilizing fuzzy logic along with mathematical process. Upon the normalized training data, FuzzITree algorithm is applied here [9]. Observations are made upon the experimental results achieved and it is seen that except for the transaction number 7, all other transactions are classified in correct manner. In order to determine the detection rate in various situations, tests are conducted on various other transactions by using various values of attributes in different situations. The detection rate is observed to be around 89% as per the test results achieved. In order to identify more optimal way for fraud detection, future research is extended in which fuzzy concept is to be utilized.

Luis Vergara, et.al (2017) proposed various methods on the basis of signal processing on graphs which helps in enhancing the performance of credit card fraud detection systems [10]. A variant of iterative amplitude adjusted Fourier transform (IAAFT) as well as other two methods known as iterative surrogate signals on graph algorithms (ISSG) have been proposed here. As per the issues that are arising related to labeling cost, algorithm testing, data confidentiality and consistent modification of patterns, a reliable augmentation of the target scarce population of frauds is necessary. By using receiver operating characteristic (ROC) curves and several key performance indicators (KPI), the feasibility of enhancing detection capabilities are evaluated using various legitimate and non-legitimate transaction ratios.

Fabrizio Carcillo, et.al (2017) presented a study through which it is seen that how the fraud detection accuracy can be enhanced through the combination of previously proposed techniques [11]. In terms of fraud detection, the exploitation or exploration of tradeoff is highlighted here. An extensive analysis if performed here by utilizing complementary machine learning techniques, on the basis of real-world dataset of millions of transactions that are provided by an industrial system. As per the outcomes achieved it is seen that a robust mechanism known as High Risk Query is achieved here which can further be utilized along with semi-supervised learning to achieve future enhancements.

Rajeshwari U, et.al (2016) proposed a mechanism to detect and prevent fraudulent transaction at the time when it is being held. The transaction will be aborted later on and alerts regarding the fraudulent transactions will be sent to the card owner [12]. The decision regarding the transaction is fraudulent or not is made on the basis of value generated for Hidden Markov Model outcome. The rate of false alarms that are achieved through comparisons amongst the actual transactions that are marked as fraud, are minimized through this system. The frauds are prevented and false alarm rates are minimized through streaming analytics. On the basis of data of genuine card holder, the model is trained and is update regularly as well.

3. RESEARCH METHODOLOGY

At first stage, a detailed literature study would be conducted on the credit card fraud detection algorithm and data classification methods; and to know their advantages and disadvantages. Literature study will lead us towards refining the structure of the proposed security solution design to overcome the shortcomings of the existing schemes, while keeping their advantages intact in order to build a robust system. Afterwards, the proposed solution will be implemented in PYTHON simulator with all essential input and output parameters. Then the implementation will undergo a thorough performance analysis and detailed comparison with the existing models. The input data acquisition is done on the historical data of the credit card for last 1 to 10 years.

The historical data of credit card spending contains the readings of the spending based listings in the day to day interval analysis, which contains the starting and closing value of each day accounted in the historical data. During the implementation, the proposed model would be designed using the ensemble of the regression models with squared distance based classification for the purpose of historical data processing, which is generally utilized for the long term prediction. In this research work, the SVM classifier is applied which can classify the input data into the normal and fraud transactions. In order to perform text categorization, popularly used mechanism is the SVM classification method which has a predictive model. The data is taken as input here and classified data is given in two categories as output. For the text corpus in which each training example belongs to one of the two classes, a model is implemented best by using SVM training algorithm. Further, by constructing N-Dimensional hyperplane, the data is partitioned into two categories. In order to separate the data,

two parallel hyper planes are generated on each side of the hyper plane. Here, the distance between the two hyper planes is maximized through the separation of hyper plane. In correspondence to the partitioning hyper plane $f(X)$ which passes across the middle of two classes and divides them, there is a linearly separable data set for which a linear classification function is created. The classification of a new data instance, X_n , is done very easily through the testing of sign of function $f(X_n)$ once the function is determined:

Where X_n belongs to a positive class if $f(X_n) > 0$

For larger distance or margin, the error of the classifier can be generalized in better way. On the high dimensional feature set, this algorithm performs well and the kernel trick is utilized for creating a new linearly separable data through the transformation of non-linearly separable data. In order to perform numerical calculations and also to calculate the regression analysis, SVM can be utilized. Further, the elements can be ranked with the help of this algorithm as well. Another benefit of SVM is that the performance of SVM on the datasets that include numerous attributes is very good even through only specific cases can be accessed for training purpose. However, during the training and testing phase of SVM, the speed and size might be the issues. Also, choosing the kernel function parameters is not an easy task and thus is a disadvantage of this approach.

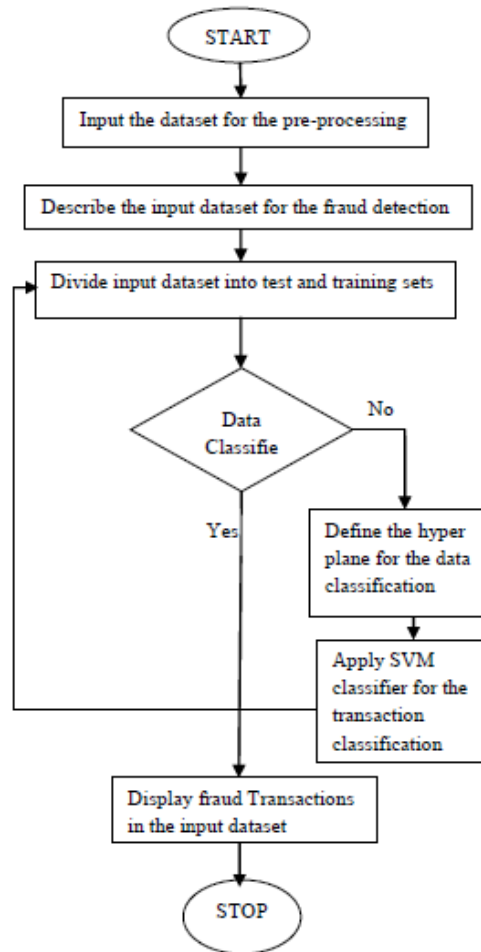


Fig 1: Proposed Flowchart

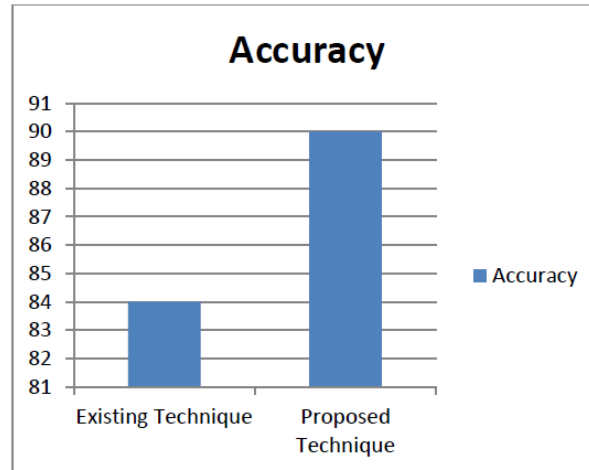


Figure 2: Accuracy Comparison

4. EXPERIMENTAL RESULTS

The proposed technique has been implemented in Python and the results have been analyzed in terms of accuracy as shown below. As shown in figure 2, the accuracy of proposed and existing algorithms is compared and it is analyzed that proposed algorithm performs well in terms of accuracy.

5. CONCLUSION

In this work, it is concluded that data mining is the technique which can mine the useful information from the large amount of data. The credit card fraud detection is the application of prediction analysis. In the prediction analysis, future data is predicted from the current information. In the existing technique, neural networks is applied which can learn from the input data and drive the future values. In this work, classification technique is applied which can classify normal and fraud transactions. The SVM classifier is applied in this work which can divided whole data into test and training sets. The test and training sets are taken as input and drive the future values. The performance of proposed modal is tested in terms of accuracy. The proposed technique shows high accuracy as compared to existing technique.

6. REFERENCES

- [1] K. C. Tan, E. J. Teoh, Q. Yu, K. C. Goh., "A hybrid evolutionary algorithm for attribute selection in data mining". Expert system with applications 2008, Elsevier
- [2] Pardeep Kumar, Nitin, Vivek Kumar Sehgal, Durg Singh Chauhan. "Selection of evolutionary approach based hybrid data mining algorithms for decision support systems an business intelligence", ICACCI ACM August 2012 Chennai, India
- [3] Mrutyunjaya Panda, Ajith Abraham, "Hybrid evolutionary algorithms for classification data mining", Springer, 10 August 2014
- [4] Rana Forsati, MohammadReza Meybodi, Mehrdad Mahdavi, AzadehGhari Neiat. "Hybridization of k means and harmony search methods for web page clustering" IEEE International Conference of Web Intelligence and Intelligent Agent Technology, 2008
- [5] Yao Yu, Fu Zhong-liang, Zhao Xiang-hui, Cheng Wen-fang. "Combining classifier based on decision tree" IEEE International Conference on Information Engineering Vol 2. July 2009
- [6] A. Shen, R. Tong, and Y Deng, "Application of classification models on credit card fraud detection", 2007, Service Systems and Service Management, 2007 International Conference on, pp. 1-4. IEEE
- [7] Kuldeep Randhawa, Chu Kiong Loo, Manjeevan Seera, Chee Peng Lim, Asoke K. Nandi, "Credit card fraud detection using AdaBoost and majority voting", 2017, IEEE
- [8] Suman Arora, Dharminder Kumar, "Selection of Optimal Credit Card Fraud Detection Models Using a Coefficient Sum Approach", International Conference on Computing, Communication and Automation (ICCCA2017)
- [9] S Md. S Askari, Md. Anwar Hussain, "Credit Card Fraud Detection Using Fuzzy ID3", International Conference on Computing, Communication and Automation (ICCCA2017)
- [10] Luis Vergara, Addison Salazar, Jordi Belda, Gonzalo Safont, Santiago Moral, Sergio Iglesias, "Signal Processing on Graphs for Improving Automatic Credit Card Fraud Detection", 2017, IEEE
- [11] Fabrizio Carcillo, Yann-Ael Le Borgne, Olivier Caelen and Gianluca Bontempi, "An Assessment of Streaming Active Learning Strategies for Real-Life CreditCard Fraud Detection", 2017 International Conference on Data Science and Advanced Analytics
- [12] Rajeshwari U, Dr B Sathish Babu, "Real-time credit card fraud detection using Streaming Analytics", 2016, IEEE