Machine Learning Approach for Credit Card Fraud Detection

Sonal
Mtech Scholar
Guru Nanak Institute of Technology
Mullana,Ambala

Mr.Kamal Gupta
Assistant professor
Guru Nanak Institute of Technology
Mullana,Ambala

Abstract

The extraction of the useful information from the raw data is done a technique known as data mining. The prediction of new things from the current data has been done using the prediction analysis which is the application of data mining. Classifications techniques are most commonly used which are implemented for the prediction analysis. Hence, prediction of the credit card fraud detection is the main objective of this work. Author proposed various credit card fraud detection mechanisms and techniques to prevent and detect fraud timely. The fundamental of the proposed technique in the base paper is based on the conventional neural networks. This system drives the new values and learns from the previous experiences. For the detection of the credit card fraud, SVM classifier is proposed in this research work using which input data is classified into normal and fraud transactions. Test and training sets are the two sub-parts of the input data. In terms of precision and recall, the normal and fraud transactions have been predicted on the basis of test and training sets.

Keywords:

Fraud Detection, SVM, KNN, Naïve bayes

Introduction

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. 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. 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 [1]. 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. 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 [2]. Similarly, objects having different properties and are dissimilar with each other are placed in separate cluster. 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 [3]. 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 [4]. 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 [5]. 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. 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 [6]. The development of new mechanism for fraud detection is becomes difficult due to the limitations of exchanging the ideas of fraud detection.

Literature Review

Kuldeep Randhawa, et.al (2017) presented the use of machine learning algorithm has been utilized for the detection of the fraud in the credit card by developing a system [7]. 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.

Suman Arora, et.al (2017) presented various issues are faced by individual while selecting appropriate fraud detection models as per done study [8]. 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.

S Md. S Askari, et.al (2017) 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 [9]. 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.

Luis Vergara, et.al (2017) 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 [10]. 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.

Fabrizio Carcillo, et.al (2017) presented for the detection of frauds and to enhance the accuracy of fraud detection ratio they combine the previously proposed techniques with new methods in this paper [11]. They mentioned the development or investigation of tradeoff, in terms of fraud detection. With the help of corresponding machine learning techniques, they performed the widespread analysis. This can be done by transacting millions of real-world dataset which has been provided by an industrial system. On the basis of performed experiments, it is concluded that High Risk Query mechanism is achieved here for the enhancement using semi-supervised leanring.

Rajeshwari U, et.al (2016) 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 [12]. 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.

Research Methodology

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 SVM

classifier in this research work. The SVM 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 SVM 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. SVM 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 SVM and it performance is good as only specific cases can be accessed for training purpose which is considered as the advantage of SVM. But, it has limitation of speed and size which occur during the training and testing phase of SVM. It is not easy to choose the parameters of kernel function hence, consider as disadvantage of algorithm.

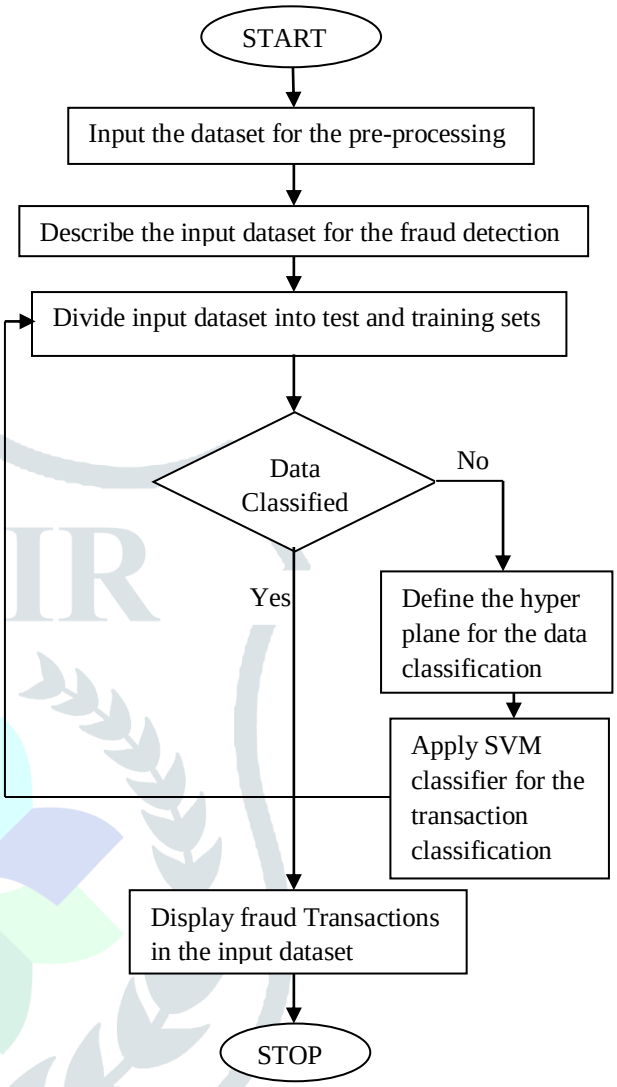

Fig 1: Proposed Flowchart

Experimental Results

The proposed technique has been implemented in Python and the results have been analyzed in terms of accuracy as shown below.

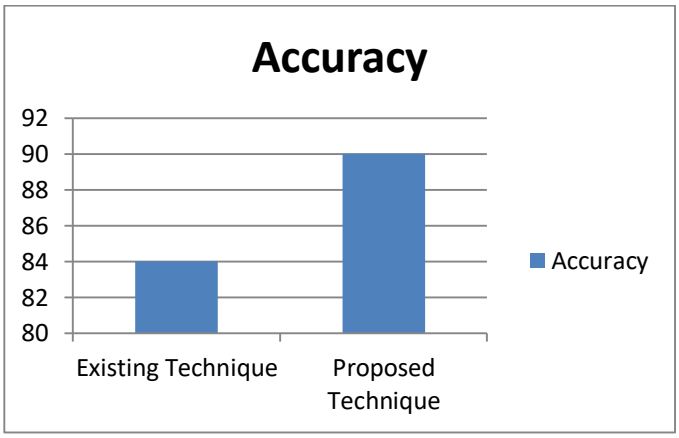

Figure 2: Accuracy Comparison

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.

Conclusion

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 SVM 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.

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, Addisson 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 Credit Card 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