A new Hybrid Approach for Credit card fraud detection using machine learning techniques

Nindiya Devi, Lavisha Sahu
Research Scholar, Center for cloud infrastructure and security Technology Suresh Gyan Vihar University, Jaipur
Assistant professor Computer engineering &Information Suresh Gyan Vihar University, Jaipur
Corresponding Author: Nindiya Devi

ABSTRACT: Protection and detection of fraud in financial transactions with credit cards at point of sale virtual represents an absolute benefit for banking, commercial and general users of these services entities. Mainly the latter two actors, are the most affected by such actions. Eradication or at least timely detection of such criminal actions would be significant progress in many areas, among which should be highlighted the growing confidence in electronic commerce. With the above, the development of a system for analyzing financial transactions with credit card arises. The system will be used by electronic commerce as a support tool for the analysis of legitimacy in financial transactions through virtual points of sale.

Keyword: Credit Card, Fraud Detection, verification, Cloud Computing

I. INTRODUCTION:

Credit cards are used as a financial tool that acts identifier and establishes responsibilities for the buyer. Using these provides credit for the purchase of goods, services and advances effective in countries around the world. Each financial institution is able to design their own credit card program to offer customers. The only thing in common between different credit programs offered by financial institutions is the establishment of credit lines between merchants and the different financial institutions worldwide. Each financial institution is able to able to set their own requirements for acceptance of credit lines and the length of periods Grace, interest on late payments, the method of calculating interest and credit limit fixed amount. The main and largest international franchises credit cards, such as Visa and MasterCard, for banks represent the first options when choosing programs that offer credit cards to their customers. Along with these are other smaller companies but which together provide a structural and operational framework for thousands of financial institutions worldwide, called affiliates, customers provide their credit card services. This framework is shaped by an international network of processing systems that allow all member institutions to authorize purchases and trade-offs between merchants and cardholders. In view of the expressed needs, developing a system for analysis of fraud with credit cards that employ techniques such as artificial neural networks, one of the branches in the field of Artificial Intelligence biggest boom in decades arises, which will allow you to learn through examples, own and different models of fraudulent transactions experiences, so that it can adapt its behavior face processing and identification of new patterns of transactions. In addition, it can be adapted to the needs of each business entity, allowing them to establish parameters to customize the type of business and to be taken into account for the best performance.

1.1 Credit Fraud Techniques

As discussed above, there are different ways in which fraudsters can carry out credit card fraud. Along with the benefits they bring advances in technology, there are new techniques and methods of fraud, which makes it increasingly difficult track these activities. Fraud can be classified broadly into 3 main categories: traditional magnetic card fraud, fraud trade body and finally Internet fraud.

1.1.1 Fraud Magnetic Card Application

This type of fraud occurs when a person falsifies an application or application to acquire a credit card. Fraud on credit applications can be carried out in 3 different ways:

1. Assume a false identity, in which an individual illegally obtained personal information of another individual and uses this information to open a credit account in your name, using partially legitimate information.
2. Financial, where the individual provides the credit issuer false information about their economic status, according to obtain credit.
3. Amount not received, also called appropriation of property postcards, where the credit card is stolen by direct mail service before it reaches the cardholder address.

II. EXITING WORK

Yukin and Zuck, 2017 [8] is a well-known one-class classification support vector machine (OCC SVM) that works with intermediate-value or set-value training data. Their original idea is to represent each distance of the training information by a limited set of obvious information with imprecise weight. Their representation is made with the expected risk of interval-value determined by the uncertain weight, which is based on the intermediate-known risk replacement generated by intermediate value information. It can be mentioned that interim anxiety is replaced with uncertain weight or possible uncertainty. The authors have shown how the limitations of unreasonable weightages are included in dual quadrilateral programming issues that can be seen as an extension of the well-known OCC SVM model. With numerical examples with synthetic and practical spacing-valuable training information, the authors decorate their proposed approach and investigate its features.

4. Miller and Soha 2015 [12] Propose novel cluster-based boosting (CBB) approach to address the limitations of the increase in supervised learning (SL) algorithms. Their CBB procedures share training data in clusters containing very similar member data and these clusters directly consolidate into the buzzing process. Because of their CBB system focusing on accurate training data, both of them try to address two specific limitations for current boosting. If the training information contains the problem area and / or the label shore, the first is filtered for the next functions; And the second is compelled to learn all the wrong instances in the next functions. The author demonstrated CBB's performance through the results of extensive experience in 20 UCI benchmark datasets and announced that CBB has achieved higher predictive accuracy using selective development without clusters.
5. Sun et al., 2016 [10] quoted a representative approach named noise-detection based AdaBoost (ND_AdaBoost) in order to improve the robustness of AdaBoost in the two-class classification scenario. In order to resolve the dilemma a robust multi-class AdaBoost algorithm (Rob_MulAda) is proposed by the authors whose key ingredients include a noise-detection based multi-class loss function and a new weight updating scheme. The authors claim that their experimental study indicates that their newly-proposed weight updating scheme is really more robust to mislabelled noises than that of ND_AdaBoost in both two-class and multi-class scenarios. As well, through the comparison experiments, the writers also verified the effectiveness of Rob_MulAda and provide a suggestion in the concrete noise level in practical applications according to the most appropriate noise alleviating approach.
6. Selma and Altay in the year 2001, an algorithm is proposed to find item sets frequently in the transaction database [27]. The basic concept of algorithm is directly inspired from the hashing and holiday (DHP) algorithm, which is a variation of the well-known Apriori algorithm. Other papers often offer a smart SMD (sorted mine) algorithm to find out the item set. This procedure, proposed by Jeba and Victor in 2011, reduces the number of database scans [28]. Based on the proposed SMINE algorithm graph construction works well. In most studies neural networks have been applied to diagnose cardiovascular disease, mainly by identifying and categorizing risky people from their ECG waveforms. In 1998, the work of cellar and chelz [29] was used to classify general and unusual ECG waveform nervous networks. Prophecy of heart disease, blood pressure and sugar with the help of neural networks proposed by Nita Guru et al. [30] In 2007, in 2008, Xianzun studied in detail, researched the integration of the neural network model for the development of data mining and data mining methods and data mining methods at the Neural Network.
7. Encryption and authentication technologies in e-commerce

8. Secure Sockets Layer (SSL)

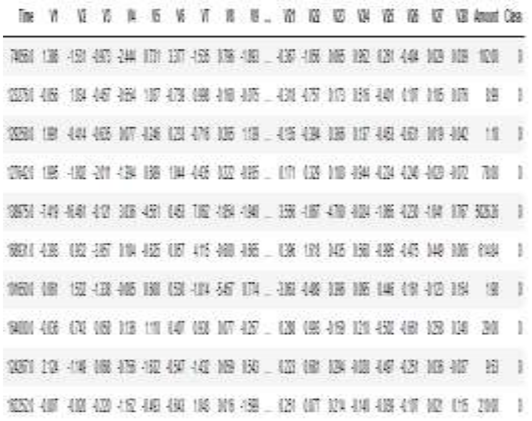
WebOpedia.com (2004) defines the Secure Sockets Layer (SSL) as a protocol developed by Netscape for transmitting private documents via the Internet. This works by using a private key to encrypt data that is transferred over the SSL connection. Both Netscape Navigator and Internet Explorer support SSL, and many Web sites use this protocol to obtain confidential user information, such as credit card numbers. By convention, the URL (see Appendix C) that require an SSL connection start with https: // instead of http: // ".

Whatis.com (2004) defines the Secure Sockets Layer (SSL) protocol commonly used as a safe operation for message transmission over the Internet. SSL uses a program layer located between the hypertext protocol (HTTP) and transport protocol (TCP). This is included as part of Microsoft and Netscape browsers, as well as most Web products. Developed by Netscape, SSL has also won the support Microsoft and other developers of client-server applications. SSL uses RSA encryption system of private and public key, which also includes the use of digital certificates. SSL is the security solution implemented in most Web servers that provide e-commerce services. His greatest merit is to answer the main problem facing online commerce: the reluctance of users to send your credit card number.

3. Proposed Work

The objective of this is to isolate the instances through random divisions of space. For this, first select a feature randomly. Then part of the domain of this variable also randomly. This process is carried out as many times as necessary until the instance is completely isolated thus generating a tree. The logic says that it would be easier to isolate the anomalous data since they will be more peculiar data and therefore, isolating these would be done with fewer separations. For this reason, the algorithm calculates an anomaly score that measures how rare this instance is. To do this, count the number of conditions that are required to isolate this instance. This is the score with which it classifies between anomalous and common instances.

4.1 Dataset Used in this Work



Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V21	V22	V23	V24	V25	V26	V27	V28	Amount	Class
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1

Fig 1: Dataset

4.2 Train and Test

This method consists of separating the data set in two, being one part (between 70 and 85%) for training and the other for testing. It should be noted that this method does not incur biases. However, a large amount of data is needed and sometimes, this can be a problem. For this data set, a separation of 80-20% was performed for the training and test sets respectively.

III. RESULT & COMPARISON

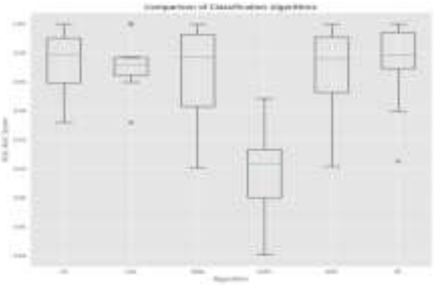


Fig 2: Comparison of Classifier

5.1 Improved K-mean:

This method of measuring the behavior of a classifier is performed by splitting the data set into k sets. Next, all but one of the k sets are used to train the model. To check the quality of the model learned, the set that has not been used to learn is used. When this is done during k iterations, k models are generated and it is ensured that all models have been trained with all possible sets and tested in the same way. In addition, one of the essential characteristics of this model is that it ensures that in each k-set, there are the same number of instances of one and another class.

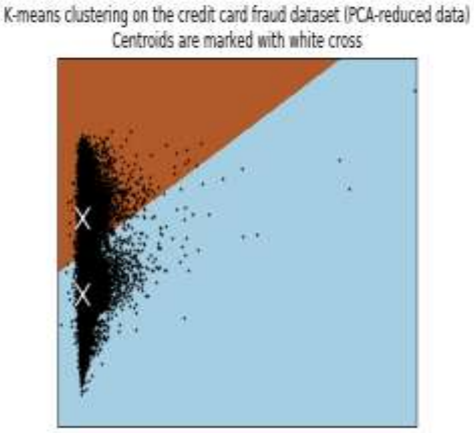


Fig 3: K- means Cluster Data classification

IV. CONCLUSION

Performing a job with such volume gives data is not a simple task, the necessary computing capacity and, therefore, the execution time, makes it often not so easy to carry out the experimentation. On the other hand, working on a real project, with data extracted directly from a public institution such as the Bank has made the motivation and the desire to obtain good results have been very high. Therefore, although at the beginning I did not understand many of the economic or tax terms necessary to understand the operation, learning these has not been a great difficulty. Taking into account the new paradigm that is introduced, one class classification, alien to what is explained in the master's classes, proves to be a very ambitious environment since it adapts very well to many of the typical problems of reality where the data is not balanced and also, it is very difficult for the anomalous event to occur. Therefore, the proposed algorithm has proven to be a very good tool by minimizing the type I error.

Future Work

To check the quality of the model learned, the set that has not been used to learn is used. When this is done during k iterations, k models are generated and it is ensured that all models have been trained with all possible sets and tested in the same way. In addition, one of the essential characteristics of this model is that it ensures that in each k-set, there are the same number of instances of one and another class. S

REFERENCES:

[1]. Cox, David Roxbee y E. Joyce Snell (1989). Analysis of binary data. Vol. 32. CRC Press. S. Zissis, D. Lekkas, "Addressing Cloud Computing Security Issues", the Proceedings of Future Generation Computer Systems, March 2012.

[2]. Armbrust M, Fox A, Griffith R, Joseph A D, Katz R, Konwinski A, Lee G, Patterson D, Rabkin A, Stoica I and Zaharia M 2010 A view of cloud computing, Communications of the ACM Magazine 53 50-58

[3]. Ashraf I 2014 An overview of service model of cloud computing Int. J. of Multidisciplinary and Current Research 2 779-783

[4]. Bala Narayada Reddy G 2013 Cloud computing-types of cloud Retrieved from

[5]. http://bigdatariding.blogspot.my/2013/10/cloud-computing-types-of-cloud.html

[6]. Christina A A 2015 Proactive measures on account hijacking in cloud computing network Asian Journal of Computer Science and Technology 4 31-34

[7]. Choubey R, Dubey R and Bhattacharjee J 2011 A survey on cloud computing security challenges and threats International Journal on Computer Science and Engineering (IJCSE) 3 1227-1231

[8]. Cloud Security Alliance 2013 The notorious nine: Cloud computing top threats in 2013 Retrieved from https://downloads.cloudsecurityalliance.org/initiatives/top_threats/The_Notorious_Nine_Cloud_Computing_Top_Threats_in_2013.pdf

[9]. Dinesha H A and Agrawal V K 2012 Multi-level authentication technique for accessing cloud services International Journal on Cloud Computing: Services and Architecture (IJCCSA) 2 31-39

[10]. Doelitzscher F, Sulistio A, Reich C, Kuijs H and Wolf D 2011 Private cloud for collaboration and e-Learning services: from IaaS to SaaS J. Computing-Cloud Computing 91 23-42

[11]. Hamlen K, Kantarcioglu M, Khan L and Thuraisingham B 2012 Security issues for cloud computing Optimizing Information Security and Advancing Privacy Assurance: New Technologies 8 150-162

[12]. King N J and Raja V T 2012 Protecting the privacy and security of sensitive customer data in the cloud Computer law & Security Review 28 308-319

[13]. Kuyoro S O, Ibikunie F and Awodele O 2011 Cloud computing security issues and challenges International Journal of Computer Networks (IJCN) 3 247-255

[14]. Li A, Yang X, Kandula S and Zhang M 2010 CloudCmp: Comparing public cloud providers Proc. of the 10th ACM SIGCOMM Conf. on Internet measurements 1-14

[15]. Malimi N 2014 Cloud computing Retrieved from http://ngeleki.blogspot.my/2014/03/what-is-cloud-computing.html

[16]. McDowell M 2009 Understanding denial-of-service attack Retrieved from https://www.us-cert.gov/ncas/tips/ST04-015

[17]. Mell P and Grance T 2011 The NIST definition of cloud computing Retrieved from http://dx.doi.org/10.6028/NIST.SP.800-145

[18]. Akhilomen, John (2013). "Data mining application for cyber credit-card fraud detection system". En: Industrial Conference on Data Mining. Springer, pp. 218-228.

[19]. Allan, Tareq y Justin Zhan (2010). "Towards fraud detection methodologies". En: Future Information Technology (FutureTech), 2010 5th International Conference on. IEEE, pp. 1-6.

[20]. Baesens, Bart, Veronique Van Vlasselaer y Wouter Verbeke (2015). Fraud analytics using descriptive, predictive, and social network techniques: a guide to data science for fraud detec- tion. John Wiley & Sons.

Nindiya Devi" A new Hybrid Approach for Credit card fraud detection using machine learning techniques" International Journal of Engineering Research and Applications (IJERA), vol. 9, no. 9, 2019, pp. 65-68