iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IV **Month of publication:** April 2018

DOI: http://doi.org/10.22214/ijraset.2018.4034

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

Online Banking Fraud Analysis and Investigation

S.K. Saravanan[1], Dr. G.N.K. Suresh Babu[2]
[1] *Assistant Professor (Sel.G), Valliammai Engineering College, Kattangulathur-603 203, Chennai.*
[2] *Associate Professor, Department Of Computer Applications, Acharya Institute Of Technology, Bangalore.*

Abstract: ***The significant growth of online banking frauds, fueled by the underground economy of malware, raised the need for effective fraud analysis systems. Unfortunately, almost all of the existing approaches adopt black box models and mechanisms that do not give any justifications to analysts. Also, the development of such methods is stifled by limited Internet banking data availability for the scientific community. In this paper we describe, a decision support system for online banking fraud analysis and investigation. Banking fraud analysis and decision support system is an effective semi-supervised approach to financial fraud and anomaly detection, using this decision support system is developed. This approach is split into two stages of development including the training phase and the runtime. During the training phase, it creates a profile for each user on the basis of its prior transactions. The training phase takes as input a series of transactions. It differentiates each user using a local, global and temporal profile. Local profiling aims at generating histograms by considering the list of transactions performed by each user. The global profiling aims at forming clusters for each user based on its prior transactions with correlated spending patterns. Temporal profiling is based on the prior transactions and it calculates the anomaly score for each user transaction. During runtime, it sorts the unlooked transactions that differ from the learned profiles. Moreover, it helps the analyst with a reason for analyzing the outcomes by aiding in his/her decision making activities.***

Keywords: Internet Banking, Bank Fraud, Fraud Detection, User profiling, Anomaly Detection, Supervised Learning, Patterns.

I. INTRODUCTION

The development of varied communication techniques, online payment transactions as well as e-commerce is spreading day by day. Moreover, the financial frauds associated with these transactions are also increasing which subsequently results in substantial financial losses every year globally.

Credit card fraud is practiced most frequently amongst the varied financial frauds due to its acceptance and widespread usage as it offers more convenience to its users. Financial institutions such as banks require more sophisticated techniques for detecting fraud. Financial frauds are often very hard to detect and analyze as the fraudulent behavior is changing, dispersed in distinct user profiles, and spread across huge imbalanced real world datasets (e.g. customer spending profiles, web logs, transaction logs).

Furthermore, customers rarely review their online banking history and hence are not able to disclose the fraudulent transactions at theright time. Accordingly, due to the wide usage of credit cards as a mode of payment for procuring goods and services, there comes a need to determine whether the transactions made through the use of a credit card is a valid transaction done by card holder or it is a fraudulent transaction done by the fraudster.

In traditional approaches, it can be figured out whether the transaction carried out is a valid transaction or a fraudulent transaction once the billing has been done. This leads to substantial financial losses. Thus, it is necessary to determine the fraudulent transactions prior to performing the billing actions.

Although fraud detection has a very long history, not much research has happened in this area. The cause is that the real world data is very hard to obtain since the financial institutions are not ready to disclose their sensitive customer transaction data due to the privacy restrictions implied by most of the financial institutions which also restricts the researchers to perform the experiments and get the outcomes.

II. OBJECTIVE OF THE PROJECT

This paper presents the techniques available for credit card fraud detection. The related work carried throughout by various researchers. To introduces the proposed methodology for the detection of credit card fraud. To discusses about the outcomes of the system. Closes out with the recommendations for future research. A challenging aspect is the abundance of users who perform few transactions, insufficient to build user profiles in a reasonable time frame. Unfortunately, none of the previous works in the area addresses this problem.

Fraud detection techniques containing transaction data are mainly split under two categories. The first category involves methods for identifying outliers in transaction data by correlation and dependence analysis. Second category which is based on the cluster analysis. These methods generally make use of clustering algorithms to aggregate transactions and recognize outlier transactions from the noted clusters. Predefined rules are commonly applied to arrange the transactions as either being fraudulent or legitimate.

III. LITERATURE POINT OF VIEW

Given the scarcity of labeled datasets, such a system must be able to work in an unsupervised or semi-supervised fashion (we can assume that no fraud exists in this dataset, as indicated by our collaborators). This conflicts with the requirement of the system being able to provide "readable" evidence to corroborate each alert. These peculiarities have remarkable implications for the typical statistical and datamining methods used in the outlier detection field. Although fraud detection has a very long history, not much research has happened in this area. The cause is that the real world data is very hard to obtain since the financial institutions are not ready to disclose their sensitive customer transaction data due to the privacy restrictions implied by most of the financial institutions which also restricts the researchers to perform the experiments and get the outcomes. Moreover, the authorities of the financial institutions change the field names so that the researchers don't get to know about the actual fields. Due to these confidential aspects of the real world dataset, fraud detection models have not been developed and described in the academic literature and very fewer models are implemented in the actual detection systems. Still there exist some of the successful applications that use different data mining techniques including self-organized maps, neural networks, artificial immune system, hidden markov models, fuzzy logic systems, conditional weighted transaction aggregation, frequent item-set mining, cryptographic algorithms, and outlier detection techniques in fraud detection.

IV. SYSTEM ANALYSIS

System is guided by an in-depth analysis of a real world dataset, which is paramount for our work and provides useful insights for future research. Therefore an existing system which unfortunately, almost all of the existing approaches adopts black box models and mechanisms that do not give any justifications to analysts. The major drawback is that it treats all the earlier transactions as equal, avoiding the continuous nature of the credit card transactions. As by overcoming such drawbacks in this paper that to indulged with some level of greater approaches to be progressed in data mining techniques to analysis the web logs (i.e.) Datasets also transaction logs of the customer. Thus the approach which is for providing decision support system for Online Banking Fraud. For all the process which are commenced by online banking as like fund transaction, mobile recharges and financial transactions. The dataset was anonymized by removing personally identifiable information, and substituting it with randomly-generated unique values to ensure our analysis could still link values that happened to be equal. The data contains customer transactions related to Bank transfers (i.e., money transfers from any account of the bank to any other account), Prepaid cards (i.e., transactions to top up credit on prepaid cards) Phone Recharges (i.e., transaction to refill prepaid cellphone accounts). Summarized as follows:

1) *CC_ASN:* the country from which the customer makes their connection, based on the Autonomous System.
2) *User-ID:* unique ID associated to a user.
3) *IBAN, IBAN_CC:* the identifier of the beneficiary account, and country.
4) *Card type* (i.e., the circuit), and number of the prepaid card.

Phone operator and number of the beneficiary of the top up.

V. DATA SET DESCRIPTON

To measure the quality of the dataset and of attributes, we make an exploratory analysis on their values, we show the results on the bank transfer data for brevity, but similar results are obtained for the other contexts.

Table -5.1 shows the distribution of the transaction amounts.

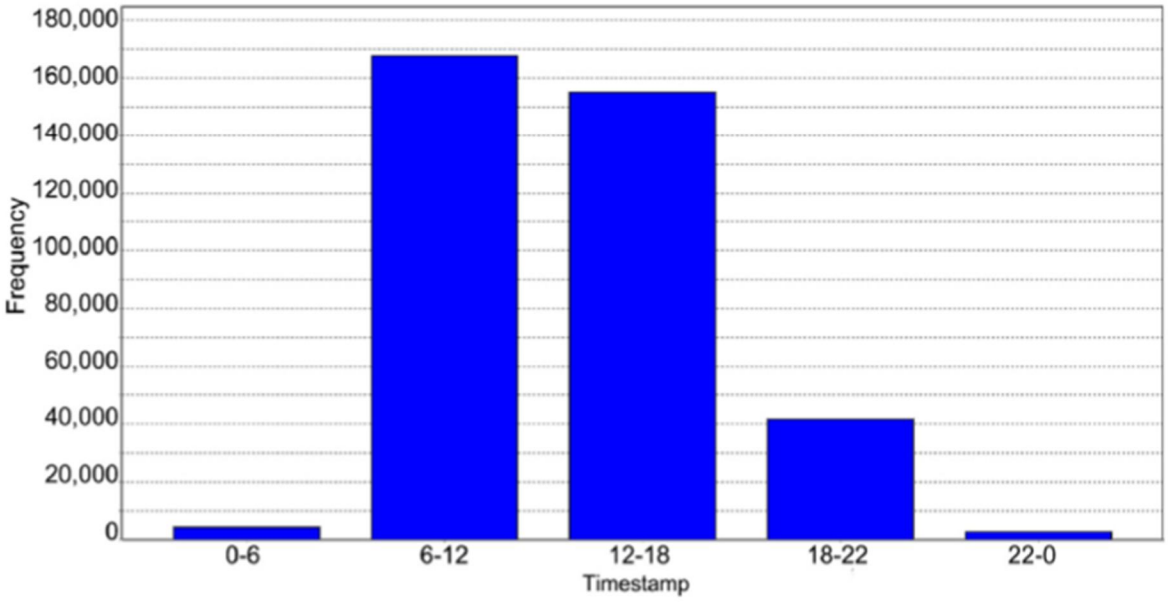
Dataset Name	Attribute Name	Description
Bank_Transfers	Amount	Non-categorical
	Account_Number	Categorical
	CC_ASN	Categorical
	Account_Number_Recipient	Categorical
	Country_Recipient	Categorical
	IP_Address	Categorical
	Time_Stamp	Categorical

VI. APPROACHES

A. Correlation Analysis

We determine to what extent features are directly correlated or dependent on each other. Attributes that share the same information (e.g., Phone operator and Phone number) and attributes derived from computations (e.g., ASN and IP) are obviously correlated. Apart from these, computing correlation on non-homogenous values requires to use approximated methods. In particular we use biserial_{rbp} methods to study the correlation between quantitative attributes (e.g., "Amount) and the categorical ones (CC_ASN, IBAN, etc.). For the correlation between categorical attributes, we use the Spearman's rank correlation coefficient, by sorting the analyzed attributes according to the frequency of each value in the dataset.

Fig 6.1.1 Discretization Number Of Transaction Per Day



In conclusion, it is not easy to estimate the dependence and the correlation between the attributes. The main obstacle is represented by the extremely sparse, imbalanced distribution of the dataset and by the high cardinality of the attributes. However, in the light of the obtained results, we decide to work under the hypothesis of independent and uncorrelated attributes. This approximation allows a much easier visualization and the interpretation of models and results, on the top a reduced temporal and spatial complexity.

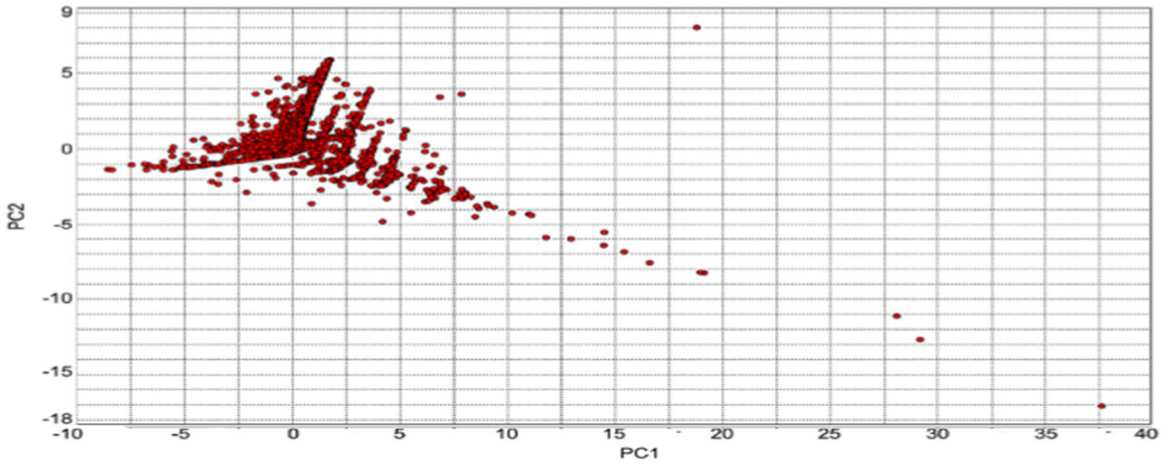
B. Cluster Analysis

We want to evaluate the feasibility of finding "classes" of users and quantifying the similarity between profiles in order to separate anomalous users from normal ones. The Principal Component Analysis (PCA) on two dimensions that we could applied to the user profiles. Instead, they tend to congregate in one dense cloud of points, with several outliers points and small groups around it.

In order to exploit the density of the large cluster, we apply DBSCAN, which is a density based clustering algorithm; it grows regions with sufficiently high density into clusters, and discovers clusters of arbitrary shape in DBSCAN algorithm which tries to separate zones with different density in the big cluster of data by executing multiple iterations of DBSCAN, using increasing values. To evaluate the quality of this clustering and to find a stopping criterion we use the Davies-Bouldin.

Index, as the number of clusters and the parameter vary. As the number of cluster grows, the index has an increasing trend, reaches a global maximum and then decreases. In the framework of online banking fraud anomaly detection that synthesizes relevant information for each user and transaction. The main objective of our system is to be a Decision Support System, able to improve the speed and accuracy of the detection of frauds by the bank analysts characterizing the local, global and the temporal profile, which are built during a training phase. In training phase takes list of transactions as input and each profile extracts with different statistical features from the transaction.

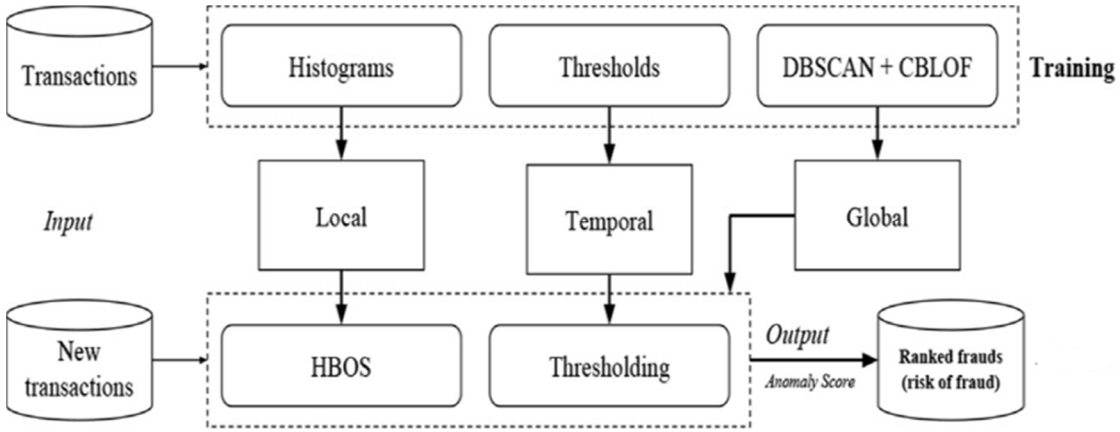
Fig 6.2.1 PCA On The Dataset On Two Dimensions



VII. SYSTEM DESIGN AND DESCRIPTION

The system description which has evaluated during the training phase as in this phase it creates a profile for each user on the basis of its prior transactions. The training phase takes as input a series of transactions. It differentiates each user by using a Local, Global and Temporal profile.

FIG-7.1-General Architecture of the System



A. Local Profile

Local Profiling illustrates foregoing user behavior to measure the anomaly of each new transaction by using "a novel algorithm" that uses the Histogram Based Outlier Score(HBOS)**.** Local profiling aims at generating histograms by considering the list of transactions performed by each user. It divides the amount for each transaction into separate ranges and provides a count of all the transaction falling into each range. This is used to generate a user details file illustrating the account number from which the transaction is made, the total number of transactions done by each account holder, and the average amount of all the transactions made. Local Profiling thus highlighting the individuals spending patterns by aggregating the transactions performed by each user.

B. Global Profile

The Global profiling clusters users based on their transaction features by means of "an iterative version of Density Based Spatial Clustering of Applications with Noise (DBSCAN)". For this, it uses "Cluster Based Local Outlier Factor (CBLOF)". Global profiling aims at forming clusters for each user based on its prior transactions. It groups the users by considering the transactions made by user based on the account number from which the transaction has taken place, the IP address of the account holder, the country from which the transaction is made, the account number to which the transaction is made and the country of the recipient's account holder.

C. Temporal Profile

Temporal profiling is based on the prior transactions and it calculates the anomaly score by employing thresholds. According to the model built, each profile generates varied statistical features from the transaction attributes. Temporal profiling calculates the anomaly score for each user as well as the anomaly score for each of the transaction made by the user. It sorts the anomaly scores in decreasing order. And the anomaly score with the highest value is considered as being abnormal and must be suspected by the analyst. The anomaly score measures the possibility of a transaction as a fraud with respect to the profiles learned. Accordingly, a general framework for semi-supervised outlier detection is developed. It employs a mixture of distinct models to discover frauds of different types. In this approach, the major aim is gathering and correctly establishing information that helps to analyze the abnormal behavior.

VIII. SYSTEM IMPLEMENTATION

A. Hidden-Morko Model

We implemented Hidden Markov Model (HMM) in the domain of credit card fraud detection. It is a sequence of real time transactions in a credit card fraud detection system. Further, they have incorporated a RFID device to display the transactions occurring over time. It observes the behavior of the customers by presenting a high security, it provides the user with the OTP(one Time Password) along with the facility to block the credit card as soon as the user realizes that the credit card is lost. It also makes sure that it does not rejects the genuine transactions by making use of the onetime password generated by the server and sent to the personal communication address of the customer (mobile phone). Have applied HMM in credit card fraud detection. This proposed divides the transaction amount into three major groups including high, medium and low transaction amounts. Each group requires different ranges of transaction amount. HMM used to represent the varied steps in a credit card transaction processing system. Further, a method has been suggested to identify the spending habits of each of the customer. This approach is scalable for managing large volumes of transactional data.

IX. RESULT

The goal is to measure the effectiveness of this system in correctly identifying the transactions that are fraudulent and are not seen before in its prior transactions. The trained dataset used consists of bank transfers, i.e. the money transfers that are carried out by an account belonging to a bank to any another account. The evaluation of this system is quite complex because, this system requires more complicated datasets that are usually very hard to obtain due to the privacy restrictions implied by the financial institutions. This work aims at developing an effective fraud analysis and automatic decision support system for banking frauds. The scores calculated have a clear statistical meaning, aiding the analyst's activity. The goal is to support the analysis of (novel) frauds and anomalies by analyzing bank transfer logs. It provides the analysts with a ranked list of fraudulent transactions, along with the anomaly score of each user sorted in decreasing order. Top-ranked transactions have higher priority. Therefore, the focus is on collecting and correctly ranking evidence that support the analysis of fraudulent behavior, rather than just flagging transactions.

X. CONCLUSION

At present, building a well-defined, manageable and well understood financial fraud monitoring system is the essential requirement of most of the financial institutions. Online Banking fraud analysis and decision support system is an effective semi-supervised approach to financial fraud detection and anomaly detection, using this decision support system is developed. Even with the typical privacy restrictions of the financial institutions it is feasible to achieve a decision support system that aids in supporting the analyst's to look over the possible reasons for the occurrence of fraud. Moreover, it helps the analyst with a reason for analyzing the outcomes making manual check much easier. Consequently, the target is on identifying the information correctly that aid in supporting the analysis of abnormal transactions.

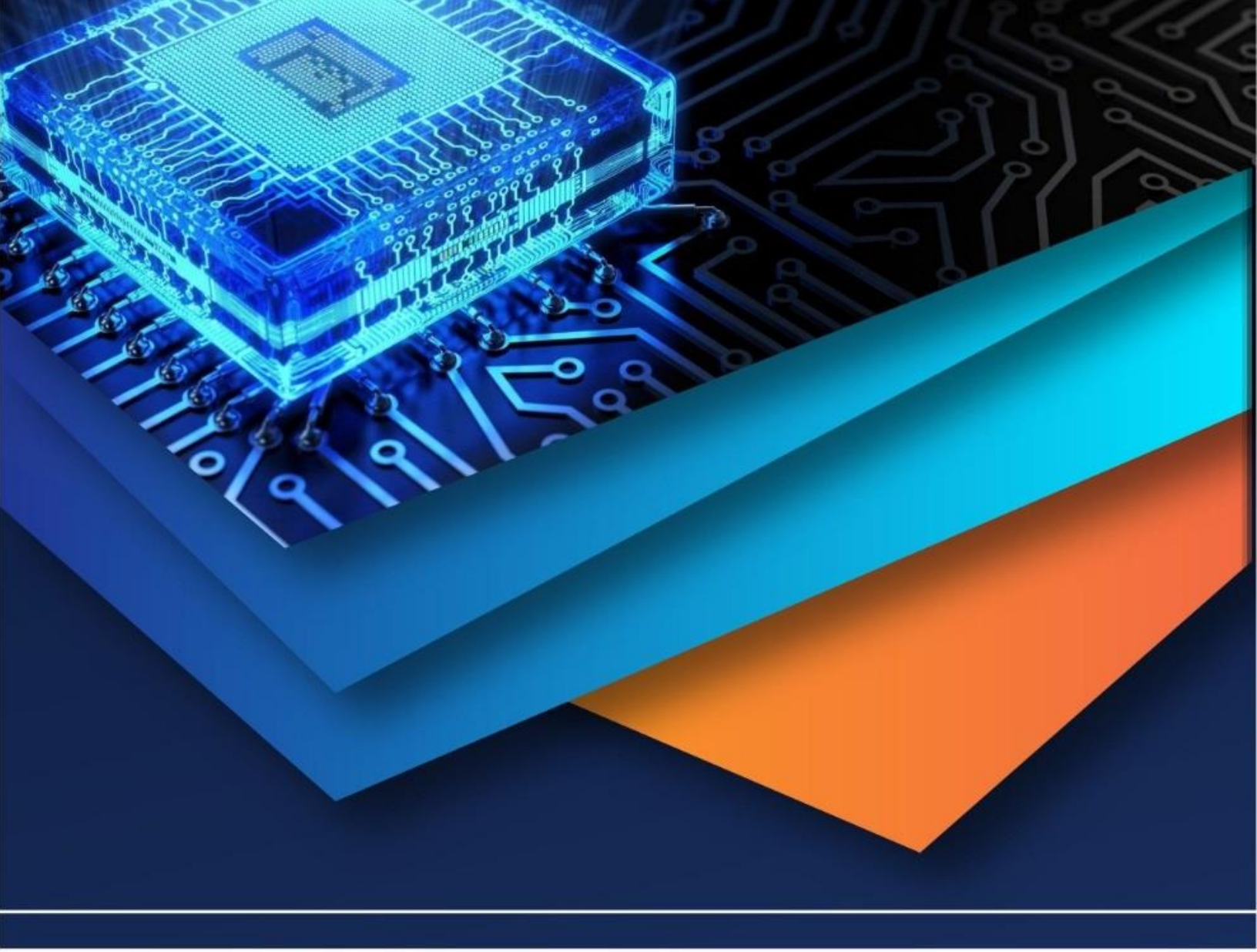
XI. FUTURE ENHANCEMENT

Future expansions are a semantic analysis of the text attributes, and a more precise estimation of the number of transactions required to fully train a profile.

REFERENCES

[1] Aggelis V. Offline internet banking fraud detection. In: ARES, IEEE Computer Society; 2006. p. 904e5.
[2] Amer M, Goldstein M. Nearest-neighbor and clustering basedanomaly detection algorithms for RapidMiner. 2012. p. 1e12.Anderson TW, Darling DA. Asymptotic theory of certain"Goodness of Fit" criteria based on stochastic processes. Ann Math Stat 1952;23(2):193e212.

[3] Wee-Yong Lim, AmitSachan, and Vrizlynn Thing, "Conditional weighted transaction aggregation for credit cardfraud detection," G. Peterson and S. Shenoi (Eds.): Advances in Digital Forensics X, IFIP AICT 433, pp. 3–16,2014, IFIP International Federation for Information Processing 2014.
[4] C. Whitrow, D. J. Hand, P. Juszczak, D. Weston and N. M. Adams," Transaction aggregation as a strategy for credit card fraud detection,"SpringerScience+Business Media, LLC 2008.
[5] MitaliBansal and Suman, "Credit card fraud detection using self organised map," International Journal of Information & Computation Technology, ISSN 0974-2239 Volume 4, Number 13 (2014), pp. 1343-1348
[6] DominikOlszewski, JanuszKacprzyk, and SlawomirZadrozny, "Employing self-organizing map for frauddetection," L. Rutkowski et al. (Eds.): ICAISC 2013, Part I, LNAI 7894, pp. 150–161, 2013, Springer-Verlag BerlinHeidelberg 2013.
[7] K. R. Seeja and MasoumehZareapoor, "FraudMiner: A novel credit card fraud detection model based on frequentitemset mining," Published 11 September 2014, The Scientific World Journal
[8] John Akhilomen, "Data mining application for cyber credit-card fraud detection system," Proceedings of the WorldCongress on Engineering 2013 Vol III, WCE 2013, July 3 - 5, 2013, London, U.K.
[9] R. Roselin and C. Hanupriya, "Customer behaviour analysis for credit card proposers based on data mining techniques," International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163Volume 1 Issue 11 (November 2014).
[10] EvaristusDidikMatyatmadja and MedianaAryuni, "Comparative study of data mining model for credit card application scoring in bank," Journal of Theoretical and Applied Information Technology, 20th January 2014. Vol.59 No.2 ISSN: 1992-8645 E-ISSN: 1817-3195.
[11] BilonikarPriya, DeokarMalvika, PuranikShweta, SonwaneNivedita and Prof.B.G.Dhake, "Survey on credit cardfraud detection using hidden markov model," International Journal of Advanced Research in Computer andCommunication Engineering, ISSN (Online) : 2278-1021 ISSN (Print) : 2319-5940 Vol. 3, Issue 5, May 2014.
[12] Ashphak P. Khan, Vinod S. Mahajan, Shehzad H. Shaikh and Akash B. Koli," Credit card fraud detection systemthrough observation probability using hidden markov model," International Journal of Thesis Projects andDissertations (IJTPD), Vol. 1, Issue 1, PP: (7-16), Month: October-December 2013.
[13] Ms. Pratiksha L. Meshram and Prof. TarunYenganti," Credit and ATM card fraud prevention using multiplecryptographic algorithm," International Journal of Advanced Research in Computer Science and SoftwareEngineering, Volume 3, Issue 8, August 2013 ISSN: 2277 128X.
[14] Ms. Amruta D. Pawar, Prof. Prakash N. Kalavadekar and Ms. Swapnali N. Tambe, "A Survey on Outlier DetectionTechniques for Credit Card Fraud Detection," IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 2, Ver. VI (Mar-Apr. 2014)



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY