iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 **Issue:** VI **Month of publication:** June 2015

DOI:

www.ijraset.com

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

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Detecting Dodgy Transactions through High Value Networks in Economic Services

Sonia[1], Anil Arora[2]
[1]M.Tech(CSE) Scholar, [2]Assistant Professor
Gateway Institute of Engineering & Technology, Sonipat

Abstract-: ***The purpose of this paper is to detect the dodgy or suspicious transactions in financial era with the help of data mining technique and also to detect more suspicious transactions in suspicious transactions. K mean technique is used to detect different clusters. Data Mining is a powerful new technology having great potential to help organizations by focusing on most important data in their corresponding data warehouse. It is the process of extracting information from large volumes of raw data. The need for data mining is because we are having large amount of data with less information in it. Data mining tools are used to analyze this data. WEKA is a data mining tool. This paper also gives use of WEKA in data mining WEKA also provides the facility for classification of data with different algorithms. Different types of bank frauds are also discussed in this paper. With the adoption of new technology as well as changing customer and staff expectations it is challenging for bank to navigate technology strategy alternatives .By using data mining, banks can differentiate their customers that what type of customers can be fraudster in future and what customers can give benefit to banks. Bank can detect highly suspicious networks from suspicious network.***

Keywords-: ***White Collar Crimes, Frauds, Data Mining, Suspicious, Clustering, WEKA Tool.***

1. INTRODUCTION

White Collar Crimes are those crimes that are committed by people of high status in the course of their occupation. Bank frauds, extortion, black mail are the different types of white collar crime. Bank Fraud involves actions in which a person is involved in the activities whose purpose is to defraud a bank of its funds. Black mail involves a person demanding money from another person using threats such as injury of property or accusation of a crime or even the exposure of a secret. Bribery is another form of white crime which involves a person giving something of value to another person with the intent of influencing their actions or persuading them to undertake certain favors.

A. White Collar Crime and Banking

Since the 1990's economic reforms, the entire banking products structure has undergone a major change. With de-regulation, increased competition and IT revolution making, it is possible to provide ease and flexibility in operations to customers, banks are also evolving and trying to become one-stop financial supermarkets. Traditionally banks were defrauded, by the fraudster by depositing stolen cheques, forged or altered cheques, fraudulent demand drafts, fraudulently procuring loans/lines of credit by submitting fake documents etc. With the advent of technology the fraudster has become more tech savvy. The focus area has shifted to technology driven products like wire transfers, internet banking, mobile banking, correspondent banking etc. Some examples of frauds in banking sector are:

1) *Cheque Fraud:* Cheques can be altered to an illegitimate payment recipient and higher transaction amount by adding a few digits or may be provided with or cheque can be make completely forged. Suspicious properties of hand or machine written cheques can be recognized by special experts [15].
2) *Loan Fraud:* Fraudulent loan applications which are reason of bank fraud may contain false information to hide financial problems. Also, an employee can knowingly approve loans to accomplices who declare bankruptcy.
3) *Money Laundering:* It is a special kind of bank fraud in which the main aim is to hide true information of origin of funds.
4) *Identity Theft:* In this fraud, the information of an individual is obtained and this information is used to apply for identity cards, accounts and credit in that person's name. The information can be obtained from mail scam, telephone.
5) *Payment Card fraud:* Payment card can be stolen or may be reproduced with skimming. Cards can be intercepted in transit

when it is being sent to the user. Card can also be negotiated by merchant who undertakes duplicate transaction of card.

II. RESEARCH BACKGROUND

The idea of this research has been taken from the book "Data Mining for Intelligence, Fraud and Criminal Detection" Advanced Analytics and Information Sharing Technologies of Christopher Westphal. Much attention has been given to financial crimes detection efforts post -9/11 era. To help combat the volume of financial crimes, a majority of international governments have created financial intelligence units to defend the integrity of worldwide financial markets .When the BSA6 was enacted, it put a mandatory requirement on banks and financial institutions, such as credit unions, savings and loans, and thrift institutions to file a Currency Transaction Report (CTR)7 for any amounts that were deposited, withdrawn, transferred, or exchanged that exceeded $10,000 in cash or coin (31 CFR 103.22). The activity has to be conducted by or on behalf of the same individual and the daily aggregate amount must exceed $10,000. Thus, if an individual went to three separate branches of a bank on the same day and deposited, say, $5,000 at each branch, the bank would be required to submit a CTR on the individual for the cumulative $15,000 deposited because it exceeds the $10,000 reporting level. CTRs are instrumental in combating all types of financial crimes and, although very powerful, their utility is somewhat limited due to certain conditions and restrictions placed on their reporting requirements. As with any system, the criminal element finds ways to circumvent the laws and new ways to launder their proceeds. Specifically, the drug dealers and organized crime members would enlist runners, mules, or smurfs to visit different banks to make deposits or purchase monetary instruments just under the $10,000 limit to avoid the filing requirements.

III. CONCEPTUAL FRAMEWORK

For solving the problem of identification of suspicious and non-suspicious transaction we have follow up the following procedure.

A. Data Mining

Data Mining automates the detection of relevant patterns in a database, using defined approaches and algorithms to look into current and historical data that can then be analyzed to predict future trends [7]. Because data mining tools predict future trends and behaviors by reading through databases for hidden patterns, they allow organizations to make proactive, knowledge-driven decisions and answer questions that were previously too time-consuming to resolve.

B. Clustering

Clustering is a data mining technique that makes meaningful or useful cluster of objects which have similar characteristics using automatic technique. The clustering technique defines the classes and puts objects in each class, while in the classification techniques, objects are assigned into predefined classes. To make the concept clearer, we can take book management in library as an example. In a library, there is a wide range of books in various topics available. The challenge is how to keep those books in a way that readers can take several books in a particular topic without hassle. In this paper, K-means clustering is used. K-means clustering is a method of vector quantization originally from signal processing, which is popular for cluster analysis in data mining. K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as prototype of the cluster [16]. This results in a partitioning of the data space into Voronoi cells. The problem is computationally difficult (NP-hard).

C. WEKA Tool

WEKA is a data mining tool. The WEKA or woodhen (Gallirallus australis) is an endemic bird of New Zealand. WEKA (Waikato Environment for Knowledge Analysis) is a popular suite of machine learning software written in Java, developed at the University of Waikato, New Zealand [15]. The WEKA suite contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality. It provides many different algorithms for data mining. GUI chooser consists of four buttons:

1) *Explorer:* An environment for exploring data with WEKA.
2) *Experimenter:* An environment for performing experiments and conducting statistical tests between learning schemes.
3) *Knowledge Flow:* This environment supports essentially the same functions as the Explorer but with a drag and drop interface. One advantage is that it supports incremental learning.

4) *Simple CLI:* Provides a simple command-line interface that allows direct execution of WEKA commands for operating systems that do not provide their own command line interface. This Java-based version (WEKA 3) is used in many different application areas, in particular for educational purposes and research. It is freely available under the GNU General Public License and is portable, since it is fully implemented in the Java programming language and thus runs on almost any architecture. It is easy to use due to its graphical user interface. WEKA supports several standard data mining tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection [15].

Figure 1: Thumbnail of WEKA

IV. PROBLEM SOLVED IN PAPER

This paper basically focuses on solving the problems of identification of suspicious and non-suspicious transactions takes place primarily in financial sector:

A. Identification of high value networks containing suspicious transactions

1) To cluster transactions into different groups based on the pattern of their profiling. Identification of high value network contains large no of suspicious transactions and further identification of clusters which have highest value in terms of no of suspicious transactions.
2) Once the cluster with highest value is identified, the high value network of transaction will further bifurcated .This process continues until we get a cluster with highest value cluster containing suspicious transactions. Now, the transactions which are included in this cluster will be identified and specified particularly.

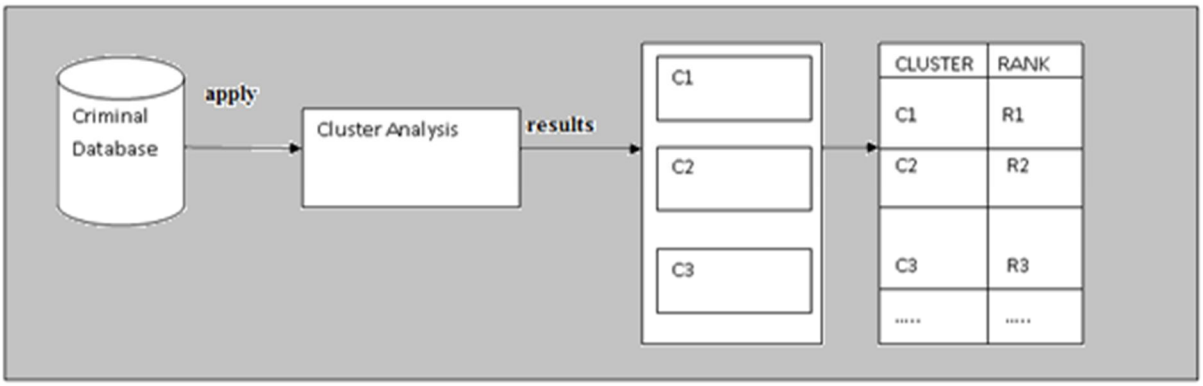


Figure 2: Description of Problem

V. DISCUSSION AND RESULT OBTAINED

The whole database will be grouped into different clusters containing suspicious and non-suspicious transactions. Among the

several clusters which have been formed by the tool on some criteria (that criteria we have considered here is total score) the network which contains the largest no suspicious transactions will be identified and explored further. For instance, if 5 clusters have been formed by the tool and have different value in terms of total no of transactions containing in those clusters then the cluster having largest value will be analyzed further classification. Real analyzation will be described further as problem is explored.

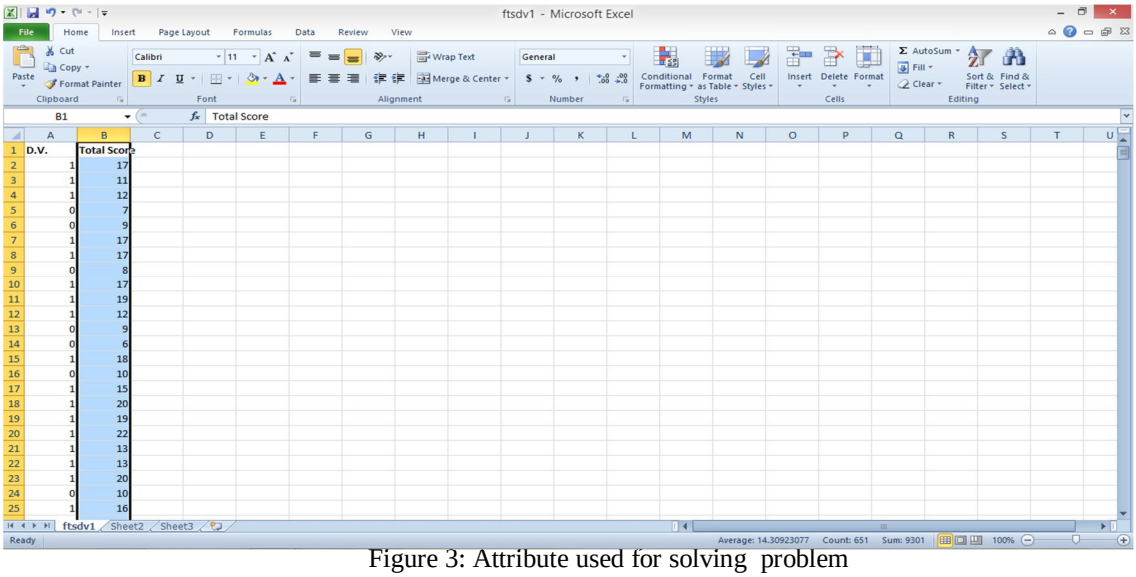


Figure 3: Attribute used for solving problem

As shown in figure 3,the attributed that will be considered for this purpose are total score and decision variable .so these two attributes will be identified from the database and copied into new excel sheet . Then store it in a .CSV file format. In Figure 4, the database we have just stored in .CSV format will be opened up and clusters will be generated automatically by the system on some criteria.

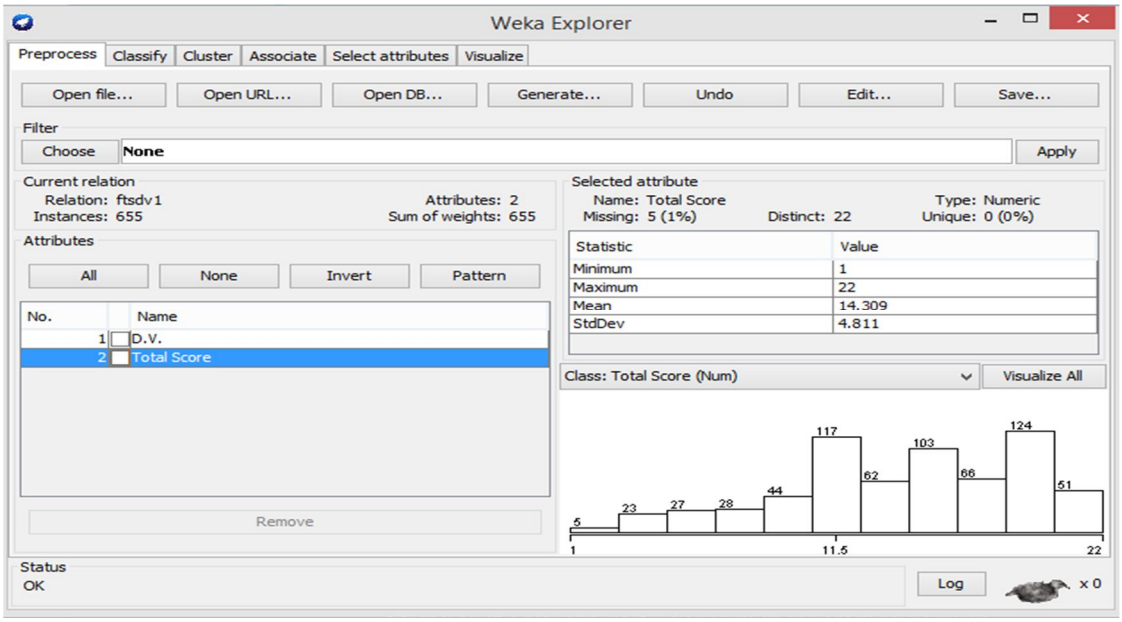


Figure 4: Results of Problem in the Form of Clusters

In figure 5, the results is obtained when clustering technique is applied as 5 clusters which having different values. For clustering, Simple K Means technique is used .This technique works as follows:

A. It identifies the clusters among the whole database according to some dynamic condition. And specify the clusters along

Volume 3 Issue VI, June 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

with the value of elements containing in a particular cluster .Also the condition on the basis of which the clusters have been formed. The total score of particular cluster is also specified.

B. Here in this figure 5, 5 clusters has been formed which contains different values .Values which defines total no of transactions suspicious or non-suspicious based on flags generated during the database formation. Here it defines in Cluster 0, there are total of 175 transactions in this cluster, and in cluster 1 there are total of 127 transactions in this cluster and so on.

C. *Cluster 0*
No of Transactions: - 175
Total Score: - 19.8571

D. *Cluster 1*
No of Transactions: - 127
Total Score: - 6.9843

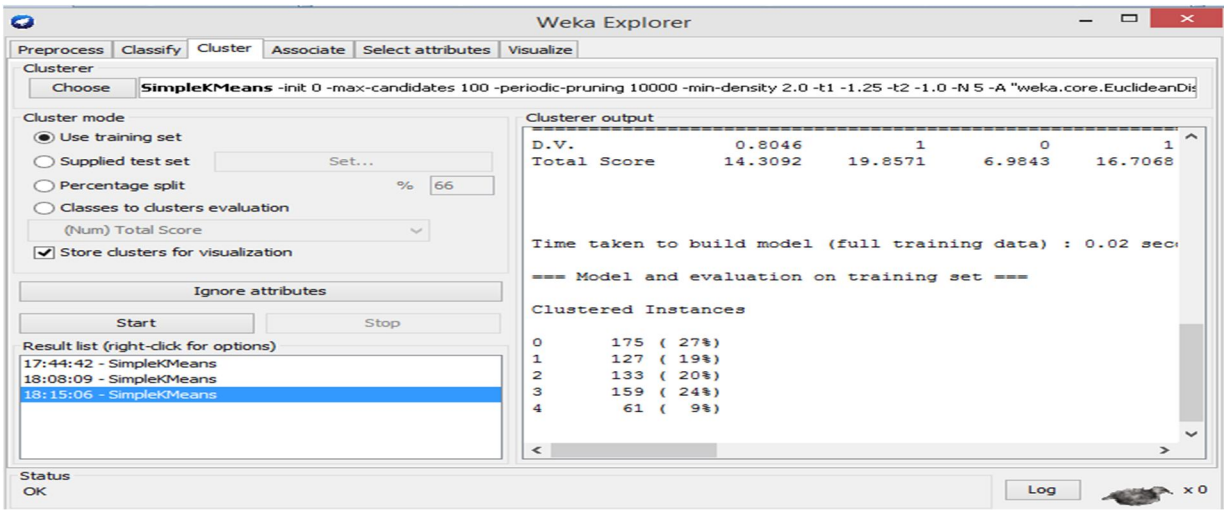
E. Cluster 2
No of Transactions: - 133
Total Score: - 16.7068

F. Cluster 3
No of Transactions: - 159
Total Score: - 11.9308

G. Cluster 4
No of Transactions: - 61
Total Score: - 14.6155

Table 1: Result 1 Obtained for the Second Problem

S.NO.	CLUSTER NUMBER	NO. OF TRANSACTIONS	TOTAL SCORE	PERCENTAGE OF TRANSACTION	RANK
1	0	175	19.8571	27%	R1
2	1	127	6.9843	19%	R4
3	2	133	16.7068	20%	R3
4	3	159	11.9308	24%	R2
5	4	61	14.6155	9%	R5



Volume 3 Issue VI, June 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Figure 5: Result After Applying Simple k means Technique

Next, the visualization of clusters which were formed in previous step will be done in this step. X: used for instance number of transaction entries in the database .Y: used for total score which has been calculated by the system for each individual cluster. Total score is the main criteria which have been used to classify the clusters in this step.

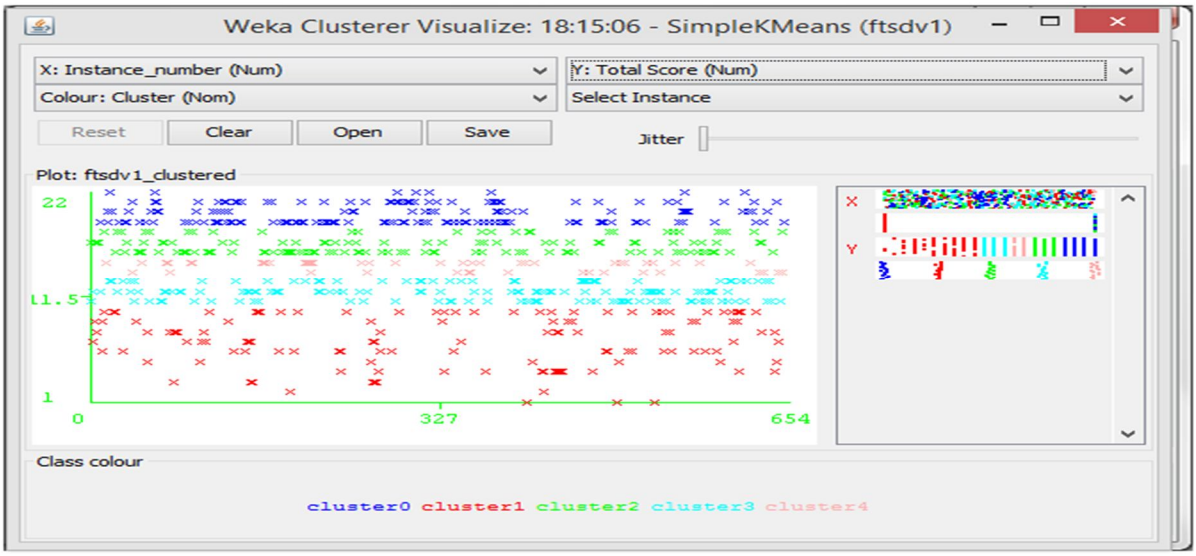


Figure 6: Cluster Shown After Applying Clustering Technique

When the total score is taken on Y-axis and D.V. is on X- axis then there will be alteration in the representation of the clusters (Figure 6). The formation of clusters will now be along the line of Y-axis ,as we have already discussed the clusters are formed on the basis of total score ,so it differentiates the clusters vertically telling that the clusters lies between different ranges of total score.

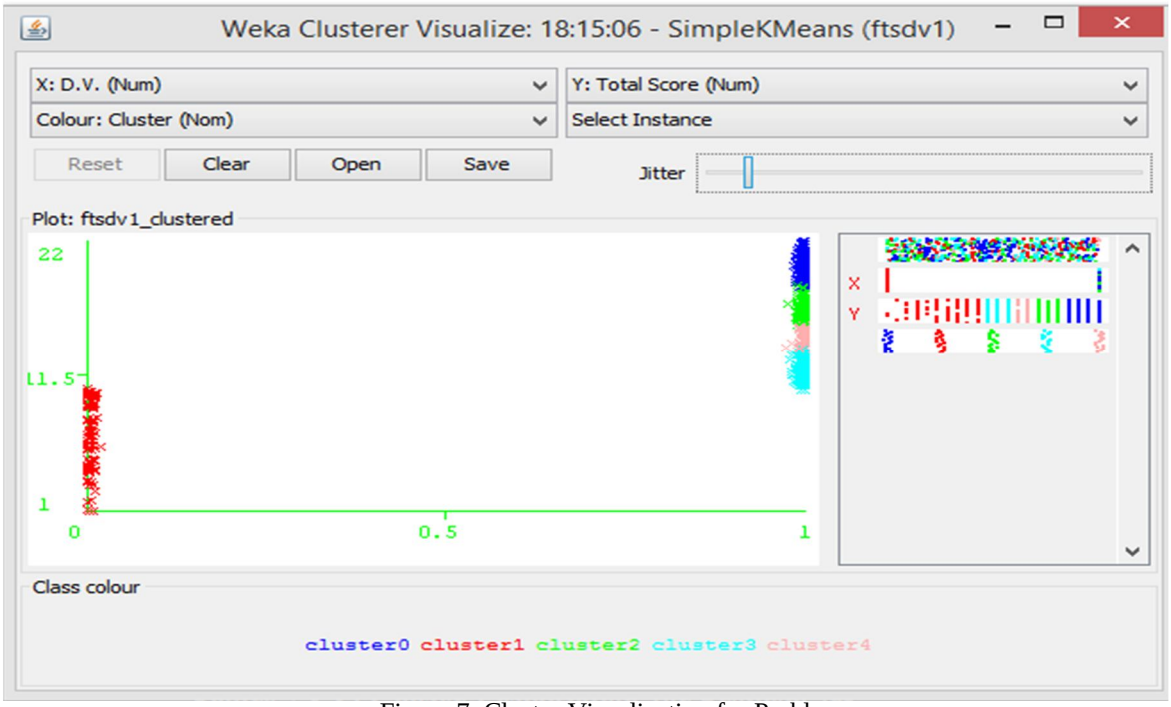


Figure 7: Cluster Visualization for Problem

A. Cluster 0 Further Bifurcated

As discussed previously, five clusters were formed after applying clustering technique. Out of these 5 clusters the cluster with highest suspicious transactions will be identified on the basis total no of transactions and total average score calculated by the system for each individual cluster. Cluster 0 is having the largest score value and total no of transactions. So this cluster will be selected to bifurcate further to detect the high value suspicious transactions. In cluster 0, there are total 175 transactions which are further bifurcated. When the cluster 0 is further divided, it breaks down into 4 clusters telling again the highest value network among these 4 networks of suspicious and non-suspicious transactions. The result obtained after applying clustering on cluster 0 is shown as below:

Cluster 0:-

No of Transactions: - 43

Total score: - 21

Cluster 1:-

No of Transactions: - 84

Total score: - 19

Cluster 2:-

No of Transactions: - 8

Total score: - 22

Cluster 3:-

No of Transactions: - 40

Total score: - 20

Table 2: Result Obtained After Cluster 0 Bifurcation

S.No.	CLUSTER NUMBER	NO. OF TRANSACTIONS	TOTAL SCORE	PERCENTAGE OF TRANSACTIONS	RANK
1	0	43	21	25%	R2
2	1	84	19	48%	R1
3	2	08	22	5%	R4
4	3	40	20	23%	R3

Volume 3 Issue VI, June 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

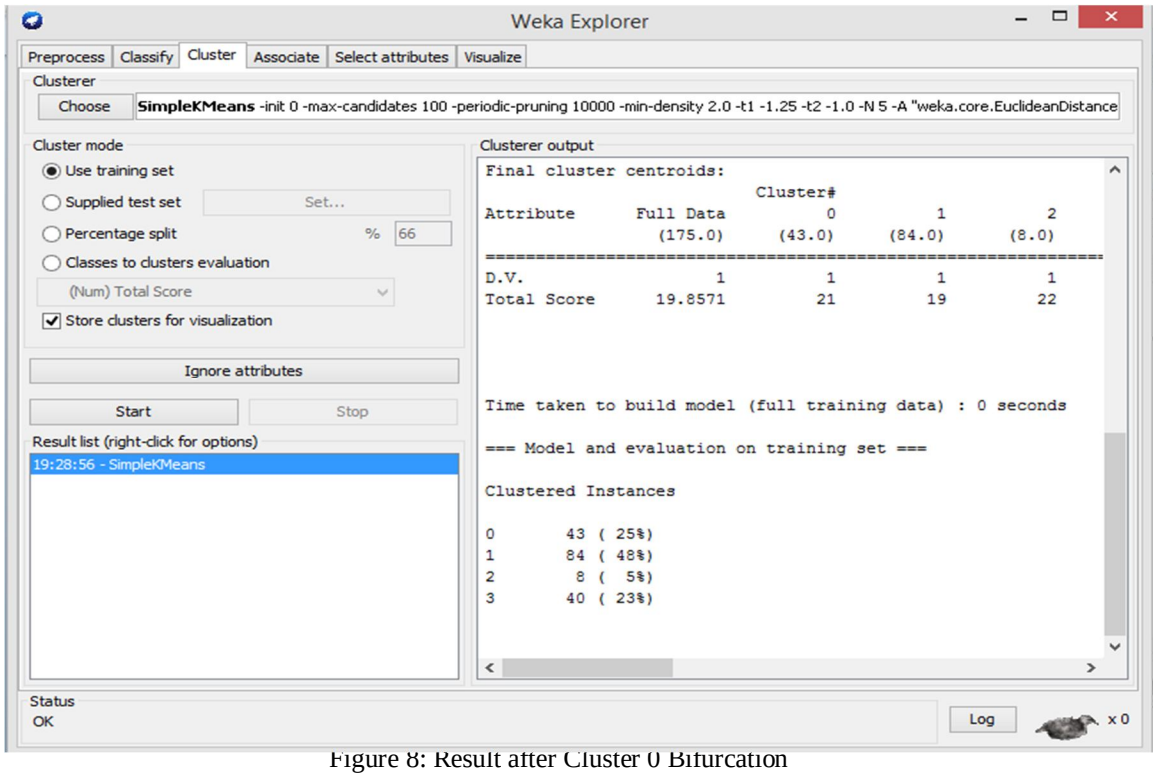


Figure 8: Result after Cluster 0 Bifurcation

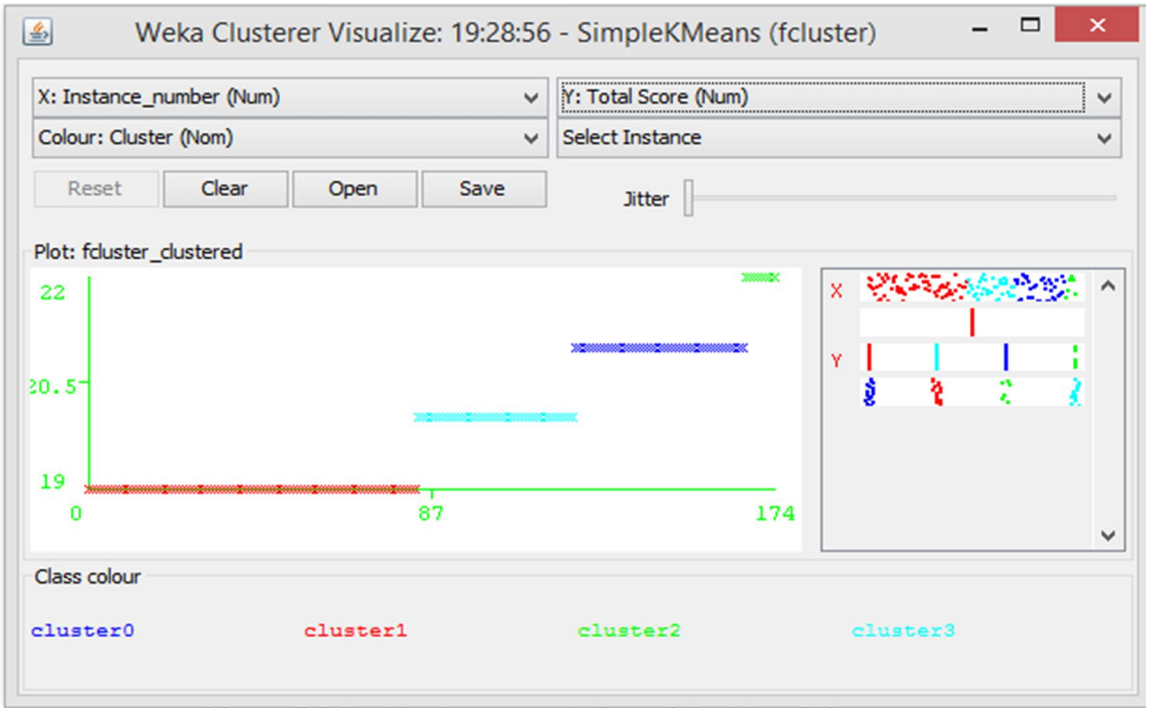
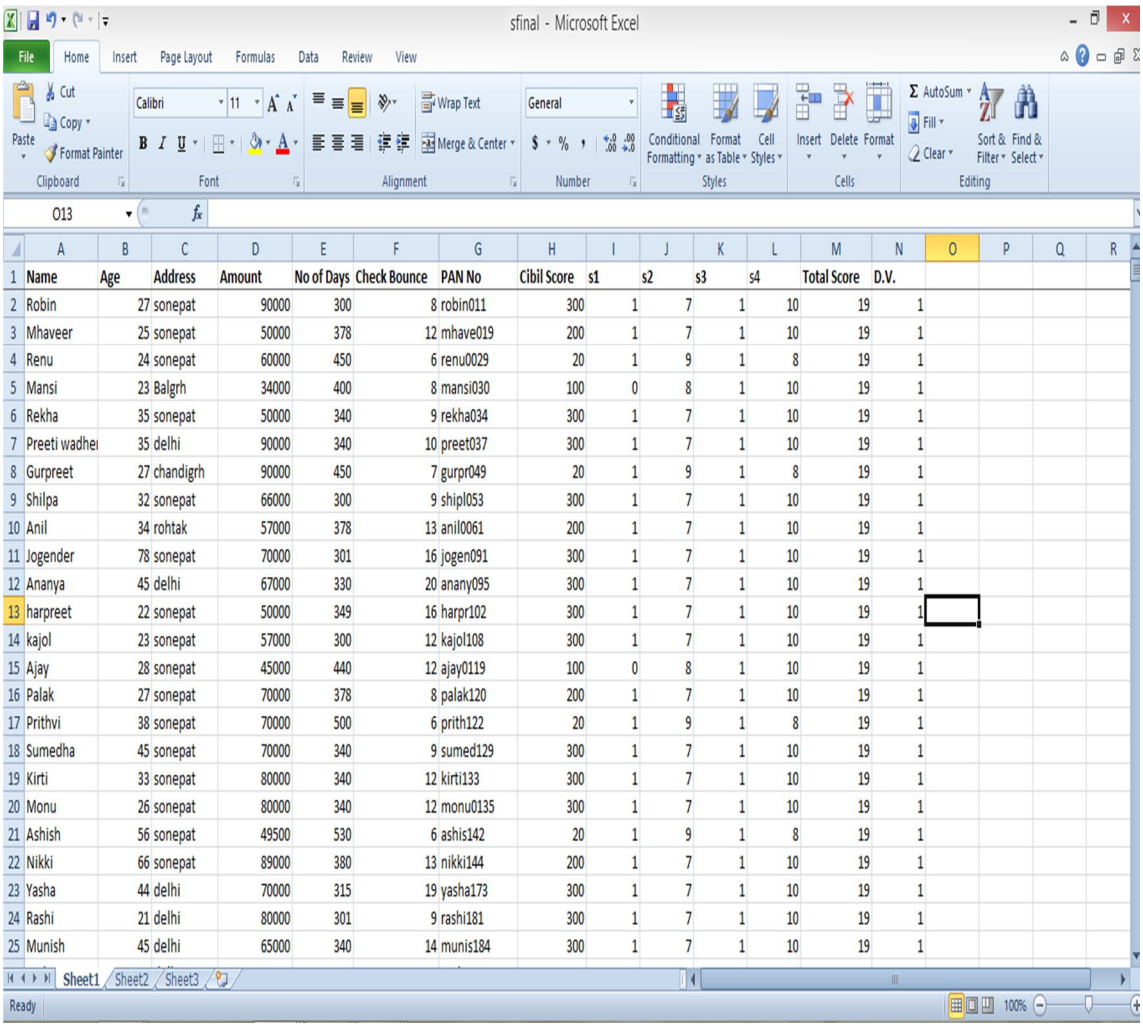


Figure 9: Visualization of Clusters after Bifurcation

B. Final Output of Problem

The final cluster i.e. cluster 2 of problem two contains the transactions which are more suspicious than other transactions. A total of 84 transactions are included in the final clusters .These transactions are shown in Figure 10. These transactions are differentiated

from other transactions by Total Score. The Total score for these transactions are defined as 19,20,21,22.



sfinal - Microsoft Excel

Name	Age	Address	Amount	No of Days	Check Bounce	PAN No	Cibil Score	s1	s2	s3	s4	Total Score	D.V.
Robin	27	sonepat	90000	300	8	robin011	300	1	7	1	10	19	1
Mhaveer	25	sonepat	50000	378	12	mhave019	200	1	7	1	10	19	1
Renu	24	sonepat	60000	450	6	renu0029	20	1	9	1	8	19	1
Mansi	23	Balgrh	34000	400	8	mansi030	100	0	8	1	10	19	1
Rekha	35	sonepat	50000	340	9	rekha034	300	1	7	1	10	19	1
Preeti wadhe	35	delhi	90000	340	10	preet037	300	1	7	1	10	19	1
Gurpreet	27	chandigrh	90000	450	7	gurpr049	20	1	9	1	8	19	1
Shilpa	32	sonepat	66000	300	9	shipl053	300	1	7	1	10	19	1
Anil	34	rohtak	57000	378	13	anil0061	200	1	7	1	10	19	1
Jogender	78	sonepat	70000	301	16	jogen091	300	1	7	1	10	19	1
Ananya	45	delhi	67000	330	20	anany095	300	1	7	1	10	19	1
harpreet	22	sonepat	50000	349	16	harpr102	300	1	7	1	10	19	1
kajol	23	sonepat	57000	300	12	kajol108	300	1	7	1	10	19	1
Ajay	28	sonepat	45000	440	12	ajay0119	100	0	8	1	10	19	1
Palak	27	sonepat	70000	378	8	palak120	200	1	7	1	10	19	1
Prithvi	38	sonepat	70000	500	6	prith122	20	1	9	1	8	19	1
Sumedha	45	sonepat	70000	340	9	sumed129	300	1	7	1	10	19	1
Kirti	33	sonepat	80000	340	12	kirti133	300	1	7	1	10	19	1
Monu	26	sonepat	80000	340	12	monu0135	300	1	7	1	10	19	1
Ashish	56	sonepat	49500	530	6	ashis142	20	1	9	1	8	19	1
Nikki	66	sonepat	89000	380	13	nikki144	200	1	7	1	10	19	1
Yasha	44	delhi	70000	315	19	yasha173	300	1	7	1	10	19	1
Rashi	21	delhi	80000	301	9	rashi181	300	1	7	1	10	19	1
Munish	45	delhi	65000	340	14	munis184	300	1	7	1	10	19	1

Figure 10: List of Transactions in Final Cluster 2

VI. CONCLUSION AND FUTURE REFERENCE

The outcomes received for the given problem helps in identification of chain of activities that contribute for occurrence of any kind of financial crime and finding out the number of transactions which are prone to major bank fraud. This problem primarily identifies the different networks which contains the suspicious and non-suspicious transactions. As mentioned above in text, the whole dataset will be divided in to different clusters. Out of these networks the network with highest value in terms of large no of suspicious transactions will be selected and investigated further to identify more suspicious networks. The bifurcation processes ends where we have got the networks with highest number of suspicious transactions and which cannot be bifurcated further on the basis of total score.

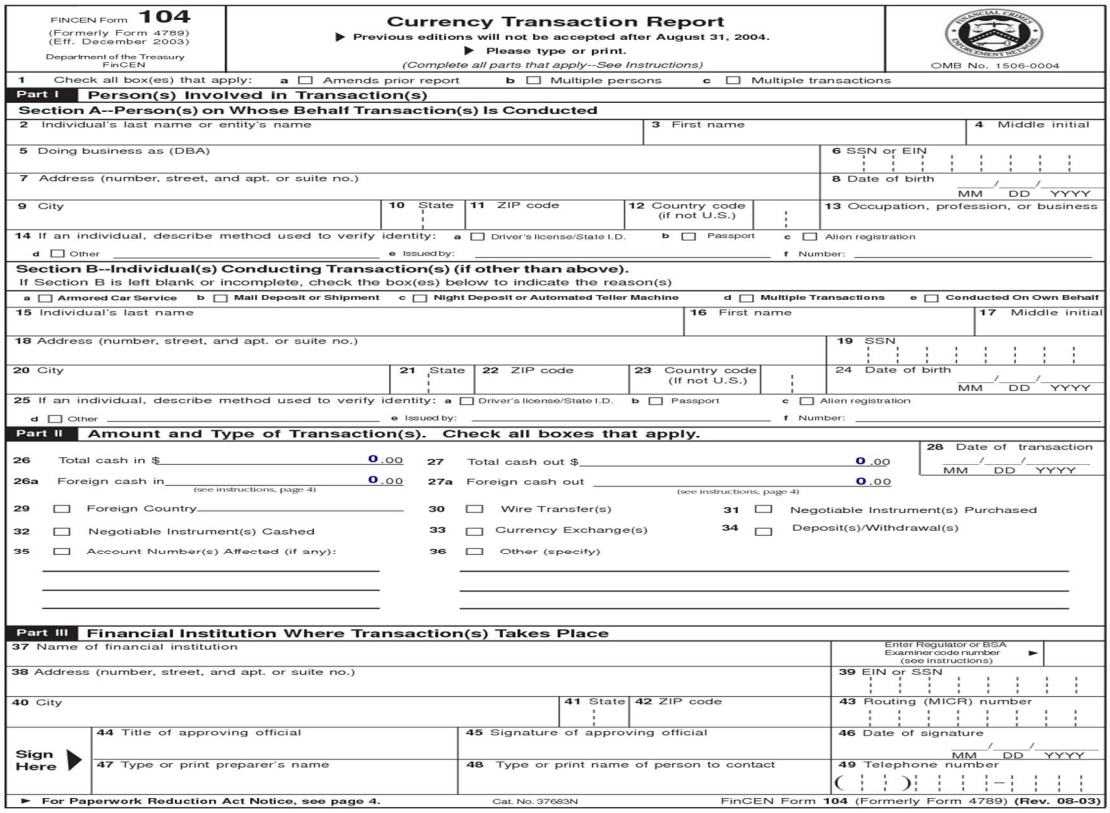
In future, this project can be useful for various reasons, some of these may be:

A. As this problem focus on detection of overall suspicious activities performed by any victim. So, in future using the patterns of these activities, the origin of crime can be identified.

B. Preventive actions to tackle these financial crimes can be taken beforehand if it is known the pattern of activities in advance.

Volume 3 Issue VI, June 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

VII. APPENDIX



FINCEN Form 104 (Formerly Form 4789) (Eff. December 2003)
Department of the Treasury FinCEN

Currency Transaction Report

▶ Previous editions will not be accepted after August 31, 2004.
▶ Please type or print.
(Complete all parts that apply--See Instructions)

OMB No. 1506-0004

1 Check all box(es) that apply: a ☐ Amends prior report b ☐ Multiple persons c ☐ Multiple transactions

Part I Person(s) Involved in Transaction(s)

Section A--Person(s) on Whose Behalf Transaction(s) Is Conducted

2 Individual's last name or entity's name | 3 First name | 4 Middle initial
5 Doing business as (DBA) | 6 SSN or EIN
7 Address (number, street, and apt. or suite no.) | 8 Date of birth MM DD YYYY
9 City | 10 State | 11 ZIP code | 12 Country code (if not U.S.) | 13 Occupation, profession, or business
14 If an individual, describe method used to verify identity: a ☐ Driver's license/State I.D. b ☐ Passport c ☐ Alien registration
d ☐ Other e Issued by: f Number:

Section B--Individual(s) Conducting Transaction(s) (if other than above).
If Section B is left blank or incomplete, check the box(es) below to indicate the reason(s)
a ☐ Armored Car Service b ☐ Mail Deposit or Shipment c ☐ Night Deposit or Automated Teller Machine d ☐ Multiple Transactions e ☐ Conducted On Own Behalf

15 Individual's last name | 16 First name | 17 Middle initial
18 Address (number, street, and apt. or suite no.) | 19 SSN
20 City | 21 State | 22 ZIP code | 23 Country code (If not U.S.) | 24 Date of birth MM DD YYYY
25 If an individual, describe method used to verify identity: a ☐ Driver's license/State I.D. b ☐ Passport c ☐ Alien registration
d ☐ Other e Issued by: f Number:

Part II Amount and Type of Transaction(s). Check all boxes that apply.

26 Total cash in $ .00 | 27 Total cash out $ .00 | 28 Date of transaction MM DD YYYY
26a Foreign cash in $ .00 (see instructions, page 4) | 27a Foreign cash out $ .00 (see instructions, page 4)
29 ☐ Foreign Country | 30 ☐ Wire Transfer(s) | 31 ☐ Negotiable Instrument(s) Purchased
32 ☐ Negotiable Instrument(s) Cashed | 33 ☐ Currency Exchange(s) | 34 ☐ Deposit(s)/Withdrawal(s)
35 ☐ Account Number(s) Affected (if any): | 36 ☐ Other (specify)

Part III Financial Institution Where Transaction(s) Takes Place

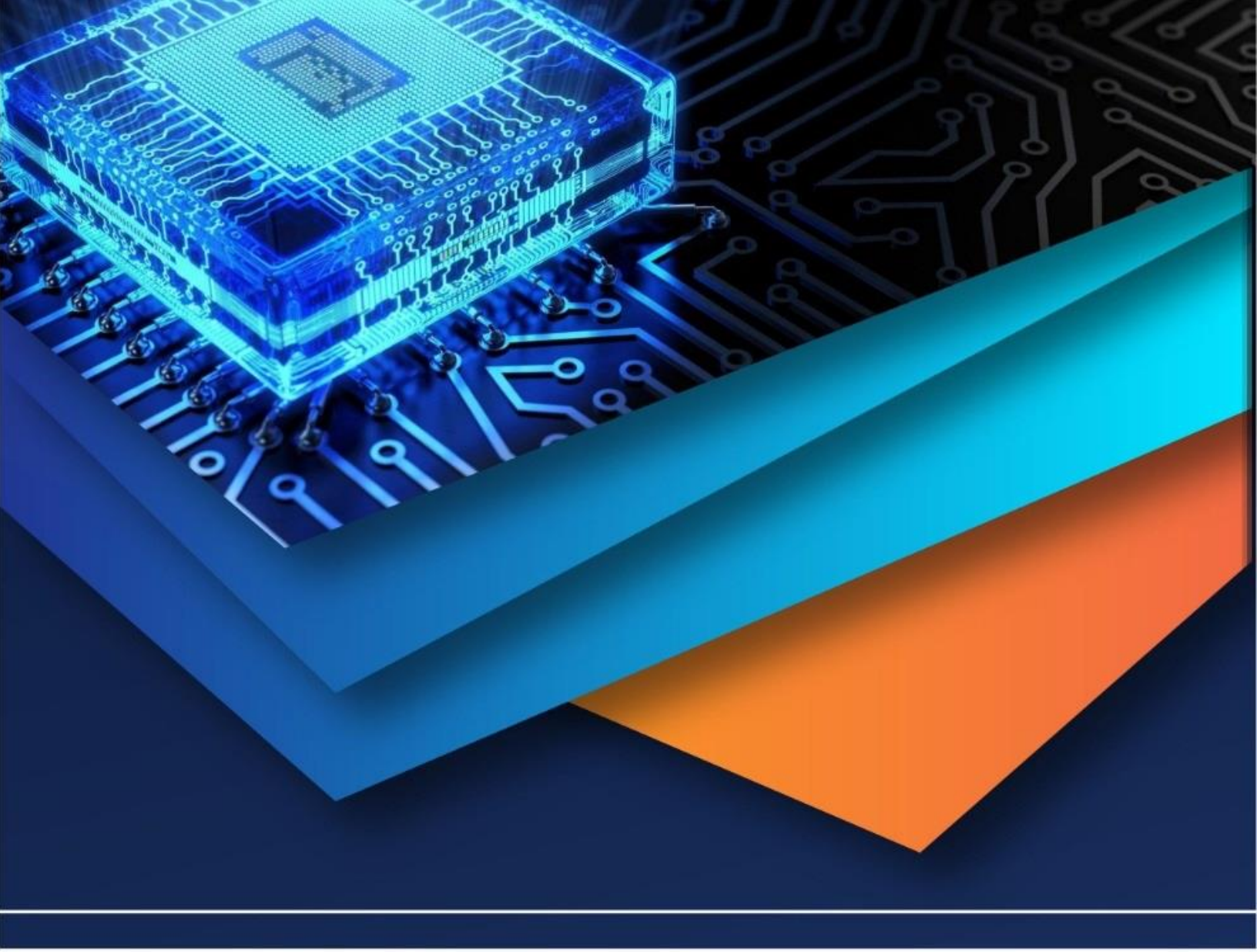
37 Name of financial institution | Enter Regulator or BSA Examiner code number (see instructions) ▶
38 Address (number, street, and apt. or suite no.) | 39 EIN or SSN
40 City | 41 State | 42 ZIP code | 43 Routing (MICR) number
Sign Here ▶ 44 Title of approving official | 45 Signature of approving official | 46 Date of signature MM DD YYYY
47 Type or print preparer's name | 48 Type or print name of person to contact | 49 Telephone number ( ) –

▶ For Paperwork Reduction Act Notice, see page 4. Cat. No. 37683N FinCEN Form 104 (Formerly Form 4789) (Rev. 08-03)

Figure 11: Currency Transaction Report

REFERENCES

[1] Jack Boorman and Stefan Ingves., Financial System Abuse, Financial Crime and Money Laundering.
[2] DP_Fraud Detection Banking. pdf, Discussion paper,2012.
[3] Financialfraud.pdf by John Howell & Co. Ltd., August 2009.
[4] JERMY QUITTNER."AVOIDING CREDIT CARD FRAUD".
[5] http://abcnews.go.com/business/_nancialSecurity/Story?id=89746andpage=12004.
[6] Abhinav Srivastava, Amlan Kundu, Shamik Sural and Arun K.Majumdar, "Credit Card Fraud Detection Using Hidden Markov Model" IEEE, Transactions On Dependable And Secure Computing, Vol. 5, No 1. , January-March 2008.
[7] M.R. Berthold et al, "Guide to Intelligent Data Analysis".
[8] IJETAE_1112_112 (1).pdf,august 2011.
[9] Financial System Abuse, Financial Crime and Money Laundering Background Paper, February 12, 2001
[10] Vol_6(3)_311 - 322_Ogwueleka, FRANCISCA NONYELUMOGWUELEKA";
[11] Trees, H.L.V. (2001). Detection, Estimation and Modulation Theory-Part I. John Wiley, New York.
[12] Stolfo, S.J.; Fan D.W.; Lee, W.; Prodromidis, A.; and Chan, P.K. (1997). Credit card fraud detection using meta-learning: Issues and initial results. Proceedings of AAAI-97 Workshop on AI Approaches to Fraud Detection and Risk Management, pp.83-90.
[13] E.Akpinar and N. Usul "Geographic Information Systems Technologies in Crime Analysis and CrimeMapping"2004.
[14] B.Muneendra Nayak et.al " A Focus on Different Frauds and Enhancing Business Process in Banking Sector Using Data Mining,", International Journal of Advanced Research in Computer Science and Software Engineering,vol.3,September 2013 .
[15] Swasti singhal et.al "A Study on WEKA Tool for Data Preprocessing, Classification and Clustering", International Journal of Innovative Technology and Exploring Engineering, pp. 250-253, vol.2, May, 2013.
[16] Ritu Chauhan et.al."Data Clustering Method for Discovering Clusters in Spatial Cancer Databases", International Journal of Computer Applications Volume 10– No.6, pp. 9-14, November, 2010.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY