ISSN:0975-9646

Investigating Suspicious Transactions through High Value Networks in Financial Intelligence Services

Anu Dahiya
Student of M.Tech (C.E)

Dr. Rajan Vohra
Head Of Department

Department of Computer Science & Engineering
P.D.M College of Engineering
Sector 3A, Sarai Aurangabad, Bhadurgarhharyana,India

***Abstract*— The given paper aims to identify the high value networks containing suspicious transactions in financial intelligence sector .This primarily focuses on the identification of network which contains the highest no of suspicious transactions than other neighbouring networks. Data Mining approach is used to identify those clusters of transactions which are more suspicious than other clusters. And also differentiating the behaviour of cluster which are highly suspicious than those less suspicious.**

Key Words: High Value Networks, Financial Crime, White collar crimes, Data mining, Clustering, Classification, Weka tool.

I. INTRODUCTION:

Financial crimes can be explained as those crimes that are committed by people of high status for the purpose of their occupation, some of the most obvious types of white collar crimes include bank frauds, bribery, measure and weight crimes, extortion, black mail, counterfeit, computer frauds and insider trading, 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.

Types of White Collar Crimes :-

Computer frauds involves a person stealing information that is contained in a computer and this mostly involves information from banks and credit card information, Counterfeiting involves the copying or imitating another person's item without the authority from the owners of the original item or copy, this happens mostly in the clothing industry and electronics. Embezz1ement of funds involves the action of a person who has been trusted with the money of an organization and he or she decides to use the money for his or her own benefit and use.

Extortion is also another form of white crime in which it involves the obtaining property illegally from another person through threats or by force, Forgery is another type of crime in which a person will use instruments such a counterfeit checks and securities in the attempt to defraud the recipient.

Insider Trading involves the use of confidential information for own benefits; most of these crimes involve the issues of shares in public corporations.

time labeling them as having the same weight as the other items.

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. LITERATURE REVIEW

Sutherland used the term "white-collar criminaloid," in reference to the "criminaloid concept" initially used by E. A. Ross (1907) in Sin and Society. Focusing on businessmen who engaged in harmful acts under the mask of respectability, Ross further wrote that the criminaloid is "society's most dangerous foe, more redoubtable by far

than the plain criminal, because he sports the livery of virtue and operates on a titanic scale." Building on these ideas, Sutherland called attention to the fact that crimes were not committed only by members of the lower class.
Sutherland's appeal to social scientists to expand their focus to include crimes by upper class offenders was both applauded and criticized. On the one hand, Sutherland was lauded for expanding the focus of the social sciences. On the other hand, the way that Sutherland defined and studied white-collar crime was widely criticized by a host of social scientists and legal experts. Much of the criticism cantered around five concerns that scholars had about Sutherland's use of the white-collar crime concept. These concerns included (1) conceptual ambiguity, (2) empirical ambiguity, (3) methodological ambiguity, (4) legal ambiguity, and (5) policy ambiguity.
Sutherland was also criticized for methodological ambiguity. He defined white-collar crime as behaviors committed by members of the upper class, but his research focused on all sorts of offenses including workplace theft, fraud by mechanics, deception by shoe sales persons, and crimes by corporations (see Robin, 1974). One might say that Sutherland committed a "bait and switch" in defining one type of crime, but actually researching another variety.
A fourth criticism of Sutherland's white-collar crime scholarship can be termed legal ambiguity. Some legal scholars contended that the concept was too sociological at the expense of legal definitions of white-collar offending (Tappan, 1947). To some, white-collar crimes should be narrowly defined to include those behaviours that are criminally illegal. Some even take it a step farther and suggest that white-collar criminals are those individuals

IV. CONCEPTUAL FRAME WORK

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

DATA MINING AND CLUSTERING METHODS

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. 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.

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.

The approach in this project is using - K MEAN and WEKA tool for clustering and revenue profiling of data.
K-MEAN : k-means clustering is a method of vector quantization originally from signal processing, that 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. This results in a partitioning of the data space into Voronoi cells. The problem is computationally difficult (NP-hard)

PAPER INTRODUCTION:

This paper basically focuses on solving the problems of identification of suspicious and non suspicious transactions takes place primarily in Financial sector:
Identification of high value networks containing suspicious transactions:
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.
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.

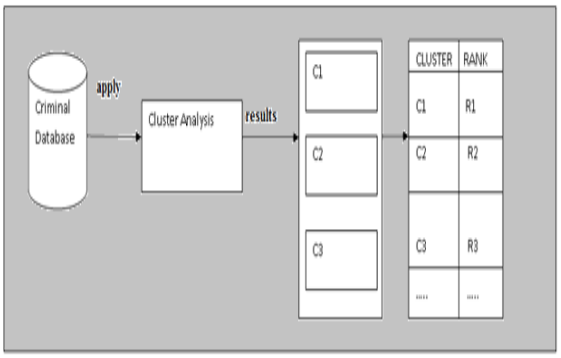

Figure 1 : Description of Problem

V. DISCUSSIONS AND RESULTS OBTAINED

The whole database will be grouped into different clusters containing suspicious and non-suspicious transactions.Among the several clusters which has 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 has been formed by the tool and has different value in terms of total no of transactions containing in those clusters then the cluster having largest value will be analysed further classification. Real analyzation will be described further as problem is explored.

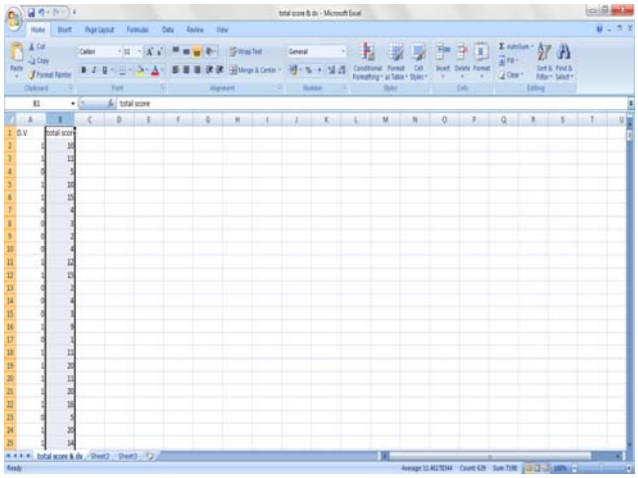

Figure 2 : Variables used solving Problem

As shown in figure 5,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 6, 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.

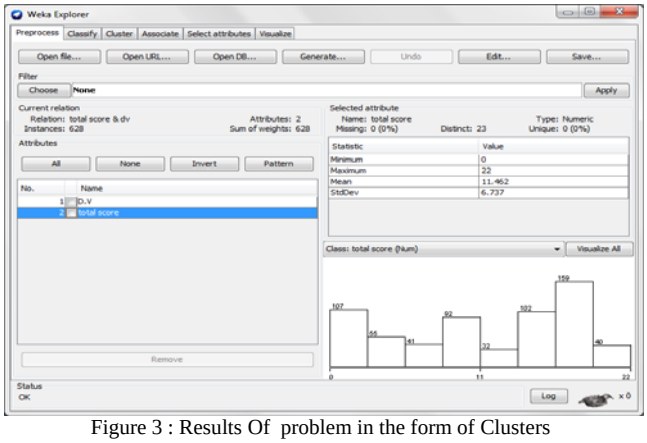

Figure 3 : Results Of problem in the form of Clusters

In figure 7 ,the results is obtained when clustering technique is applied as 5 clusters which having different values

For clustering SimpleKMeans technique is used .This technique works as follows:

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

Here in this figure, 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 152 transactions in this clusters ,in cluster 1 there are total of 149 transactions in this clusters and so on.

Cluster 0 :-
No of Transactions: - 152
Total Score: - 19.074
Cluster 1:-
No of Transactions: - 107
Total Score: - 19.074
Cluster 2:-
No of Transactions: - 149
Total Score: - 15.919
Cluster 3:-
No of Transactions :- 74
Total Score :- 4.473
Cluster 4:-
No of Transactions :- 146
Total Score :- 10.321

S.no.	Cluster number	No of transactions	Total score	Percentage of transactions	Rank
1	0	152	19.0724	24%	R1
2	1	107	0.8318	17%	R4
3	2	149	15.9195	24%	R2
4	3	74	4.473	12%	R5
5	4	146	10.3219	23%	R3

Table 1 : Result 1 obtained for the problem

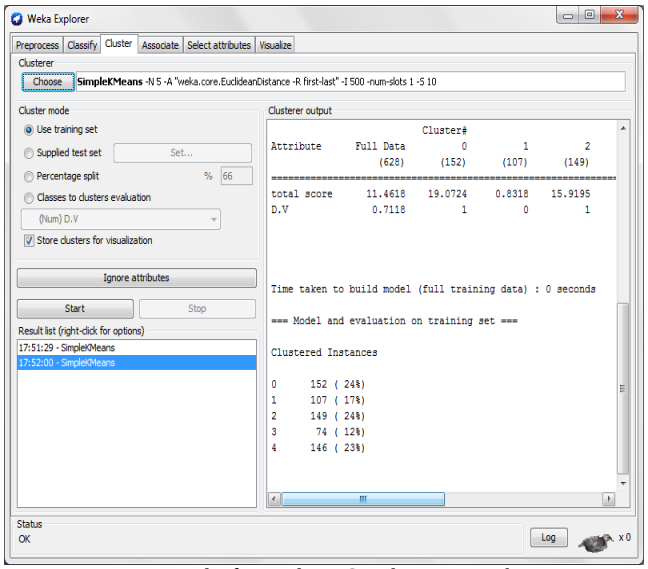

Figure 4:-Result after applying SimpleKmean Technique

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.

For example the cluster 4 contains the transaction having score value in between 8 to 12.also the total average score is calculated and displayed along with the final result of clusters formed previously.

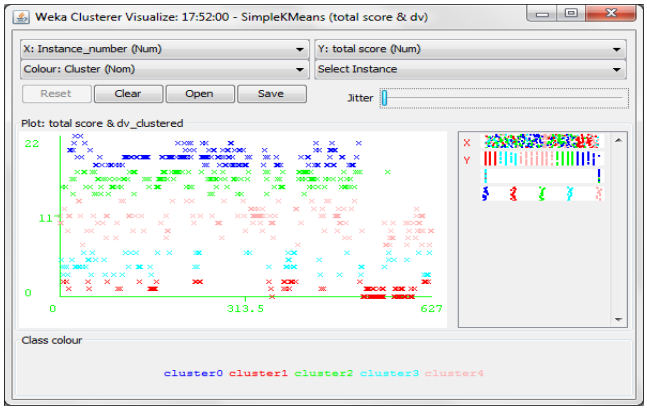

Figure 5 : Cluster shown after applying clustering technique

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. 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.

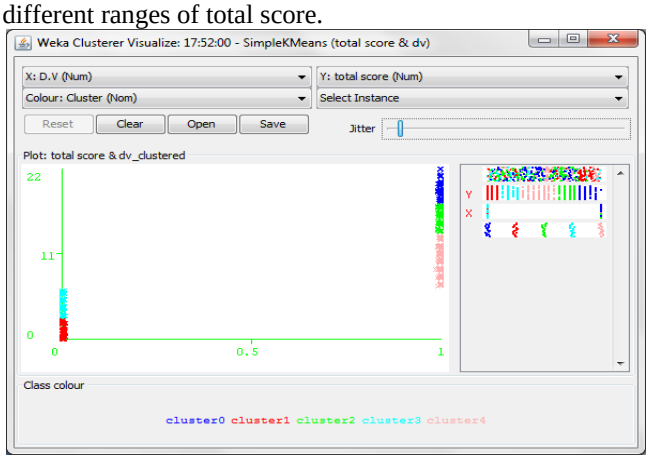

Figure 6 : Cluster Visualization for Problem

Again after changing the alignment and description along the axis there will little difference on representation of the clusters. A visibly single inclined line telling which cluster lies in which range. Range is specified horizontally and vertically.

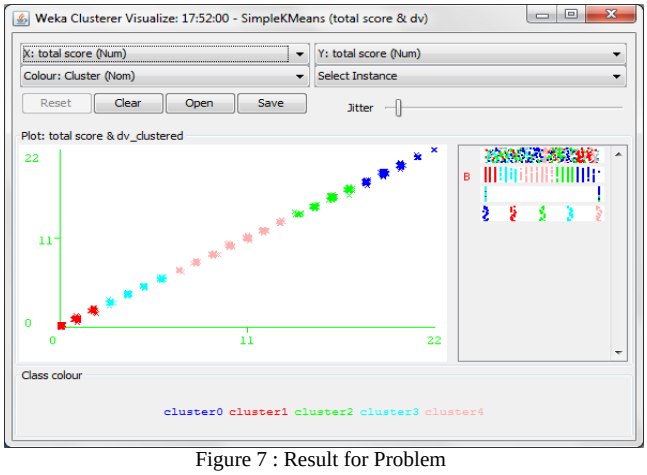

Figure 7 : Result for Problem

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.

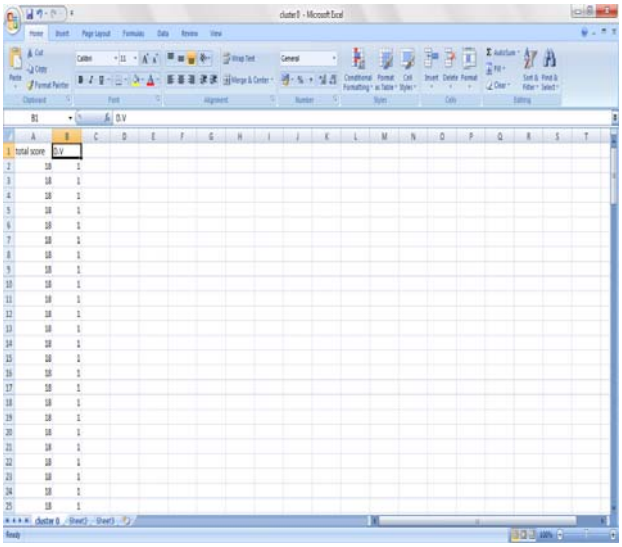

Figure 8 : List of Transactions after Cluster 0 Bifurcation

When the cluster 0 is further divided ,it breaks down into 5 clusters telling again the highest value network among these 5 networks of suspicious and non suspicious transactions.The result obtained after applying clustering on cluster 0 is shown as below :

Cluster 0 contains 45 transation whose D.V. is 1 for these transaction and taking 30% of total data and Total Score value for this cluster is calculated as 18.Cluster 1 contains 44% and having 66 transactions in total and total average score is calculated as 19 and having the highest no of transactions which will be considered for analysation further.cluster 2 contains 25 no of total transactions and 17 % of the total data base and the total average score is calculated as 20.Cluster 3 contaon only 3 tarsactions and Cluster 4 contains only 2 transactions and having the lowest no of transactions.-

CLUSTER 2 FURTHER BIFURCATED :-

Cluster 0 :-
No of Transactions :- 45
Total score :- 18

Cluster 1 :-
No of Transactions :- 25
Total score :- 20

Cluster 2 :-
No of Transactions :- 66
Total score :- 19

Cluster 3 :-
No of Transactions :- 13
Total score :- 21

Cluster 4 :-
No of Transactions :- 2
Total score :- 22

S.no.	Cluster number	No of transactions	Total score	Percentage of transactions	Rank
1	0	45	18	30%	R2
2	1	25	20	16%	R3
3	2	66	19	43%	R1
4	3	13	21	9%	R4
5	4	2	22	1%	R5

Table 5.2: Result obtained for the 2nd problem

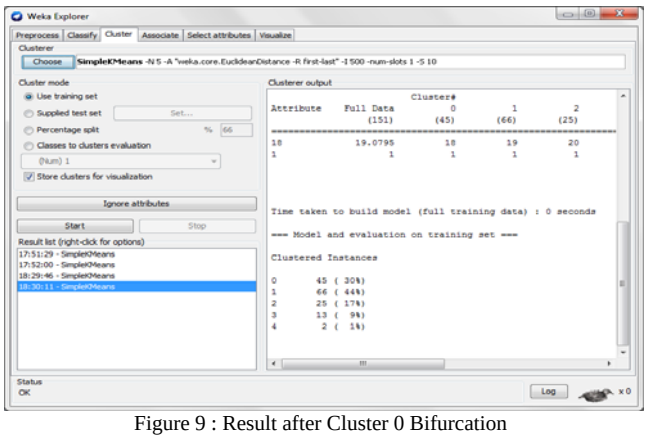

Figure 9 : Result after Cluster 0 Bifurcation

FINAL RESULT OF PROBLEM 2nd :-

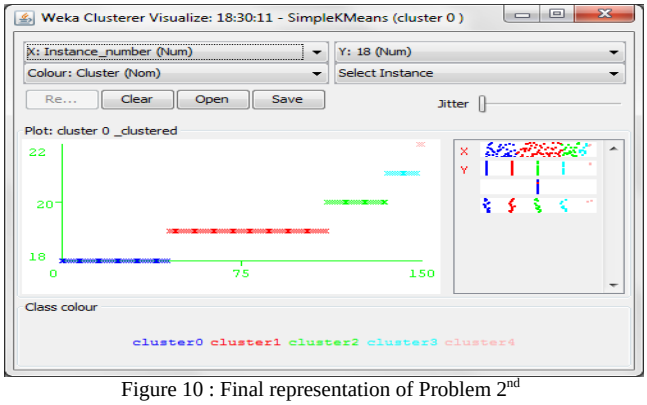

Figure 10 : Final representation of Problem 2nd

FINAL OUTPUT OF PROBLEM 2nd :-

The final cluster i.e. cluster 2 of problem two contains the transactions which are more suspicious than other transactions . A total of 66 transactions are included in the final clusters .These transactions are shown in Figure 18. These transactions are differentiated from other transactions by Total Score . The Total score for these transactions are defined as 19.

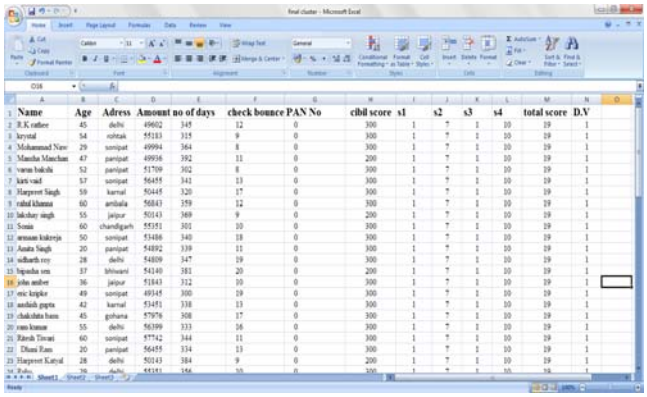

Figure 11. List of Transactions in Final Cluster 2

VI. CONCLUSION AND FUTURE WORK

The outcomes received for the given problem helps in identification of chain of activities that contribute for occurrence of any kind of financial crime. No 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 netwoks with highest no 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:

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

VII. APPENDIX

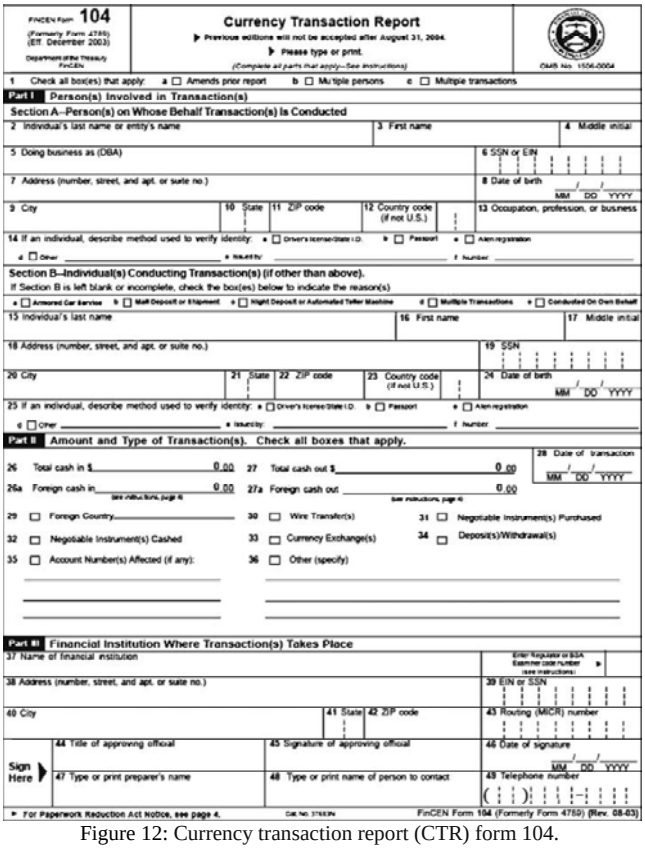

Figure 12: Currency transaction report (CTR) form 104.

ACKNOWLEDGEMENTS:

Author would like to thanks to her head Dr. Rajan Vohra, HOD of CSE & I.T department, PDMCE, Bahadurgarh for his valuable support and help.

REFERENCES:

1. http://www.thearling.com/text/dmtechniques/dmtechniques.htm
2. Jack Boorman and Stefan Ingves., Financial System Abuse, Financial Crime and Money Laundering
3. DP_FraudDetectionBanking.pdf,Discussion paper,2012
4. Financialfraud.pdf by John Howell & Co. Ltd.,August 2009
5. JERMY QUITTNER."AVOIDING CREDIT CARD FRAUD".
6. http://abcnews.go.com/business/_nancialSecurity/Story?id=89746andpage=12004
7. Abhinav Srivastava, Amlan Kundu, Shamik Sural and Arun K. Majumdar, "CreditCard Fraud Detection Using Hidden Markov Model" IEEE, Transactions On Dependable And Secure Computing, Vol. 5, No 1. , January-March 2008
8. M.R. Berthold et al, “Guide to Intelligent Data Analysis”;
9. IJETAE_1112_112 (1).pdf,august 2011
10. Financial System Abuse, Financial Crime and Money Laundering Background Paper, February 12, 2001
11. file:///material/Thesis%20papers/Pragmatic%20Programming%20Techniques%20%20Fraud%20Detection%20Methods.htm
12. Vol_6(3)_311 - 322_Ogwueleka, FRANCISCA NONYELUM OGWUELEKA”;
13. Trees, H.L.V. (2001). Detection, Estimation and Modulation Theory-Part I. John Wiley, New York.
14. 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, 83-90.
15. E.Akpinar and N. Usul “Geographic Information Systems Technologies in Crime Analysis and CrimeMapping”2004