International Journal of Research in Advent Technology (E-ISSN: 2321-9637) Special Issue
1st International Conference on Advent Trends in Engineering, Science and Technology
"ICATEST 2015", 08 March 2015

Discovering Inception of Financial Fruads

Anu Dahiya
Assistant Professor,Hindu Institute Of Management,Sonepat,Haryana-131001
Address : 275,Sector-23,Sonepat,Haryana-131001
Email Id: anu.dahiya71@gmail.com ,Phone No: 9729687231

Abstract :- This paper aims to identify the most suspicous source of white collar crimes among the several origins.We have detected by analying the data that the crimes occuring are more severe in metro cities rather than small cities. Because no of literate crimals tends to be active more in fast cities.The data to identify the source has been collected and simulated from a financial institution.After collection of data set the EM clutering technique has been applied on the data .

KeyWords:-White-CollarCrimes,Fiscal Crimes,Weka Tool,Clustering,EM.

1. 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.

2. 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.

3. Conceptual Framework :-

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

1. First the data set for financial crime has been collected and simulated.
2. Weka tool has been used to preprocess the data and EM clustering technique has been used for clustering most suspicious sources in terms of their origin.
3. EM identifies the clusters and rank according to their severity
4. After this the clusters has been visualized to identify their exact source.

3.1 EM Algorithm:-

Cluster analysis (or clustering) is the classification of objects into different groups, or more precisely, the partitioning of a data set into subsets (clusters or classes), so that the data in each subset (ideally) share some common trait - often proximity according to some defined distance measure. Data clustering is a common technique for statistical data analysis, which is used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics. The computational task of classifying the data set into k clusters is often referred to as k-clustering.

The EM algorithm implemented in BEAM can be regarded as a generalization of the k-means algorithm. The main differences are:

1. Pixels are not assigned to clusters. The membership of each pixel to a cluster is defined by a (posterior) probability. For each pixel, there are as many (posterior) probability values as there are clusters and for each pixel the sum of (posterior) probability values is equal to unity.
2. Clusters are defined by a prior probability, a cluster center, and a cluster covariance matrix. Cluster centers and covariance matrixes determine a Mahalanobis distance between a cluster center and a pixel.
3. For each cluster a pixel likelihood function is defined as a normalized Gaussian function of the Mahalanobis distance between cluster center and pixels.
4. Posterior cluster probabilities as well as cluster centers and covariance matrixes and are recalculated iteratively. In the E-step, for each cluster, the cluster prior and posterior probabilities are recalculated. In the M-step all cluster centers and covariance matrixes are recalculated from the updated posteriors, so that the resulting data likelihood function is maximized.
5. When the iteration is completed, each pixel is assigned to the cluster where the posterior probability is maximal.

4. Discussions And Results Obtained :-

The database of 447 record has been collected from an financial institutions and simulated. After this the records are fed to the tool which has been used for creation of clusters.

The attributes Address,Dicision Variable and total score has been used to identify the transactions which are responsible to be believed as suspicious. Also ,these attributes are fed to the EM-Algorithm for identification most suspicious cities in terms of financila crimes occuring in country.

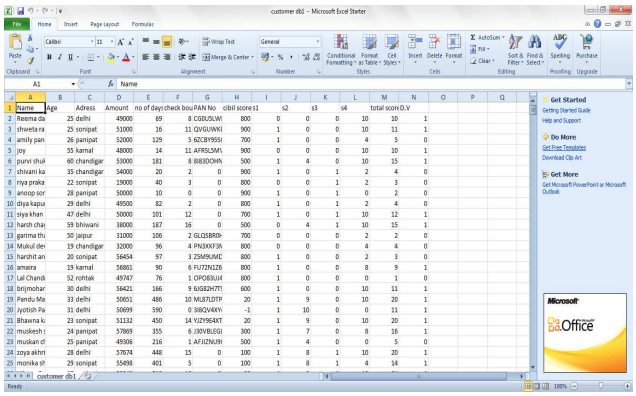

Figure 1: Database of 447 records

The database includes all the information of a financial fraud criminal and information regular visitor of a finacial organization.Which further includes his/he age,name,address,amount,no of days,PAN No,Cibil score and individual score values designated by s1,s2,s3,s4.These individual score helps to claculate our final score and further our final score helps to calculate decision variable.

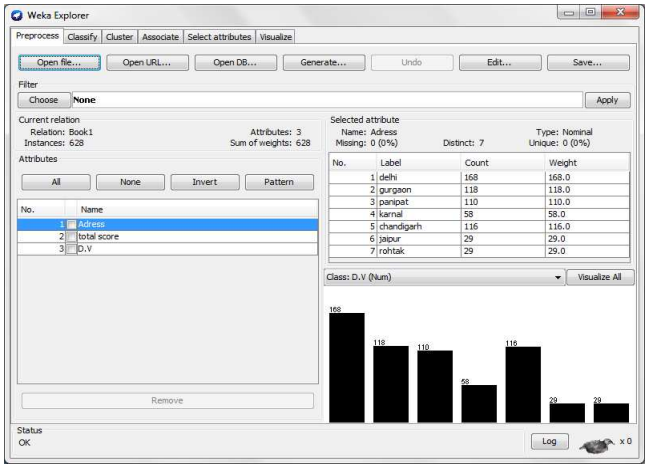

Figure 2: Preprocessed data

There are total of 10 parameters ,which are showing the profiles of criminals. But the main parameters which are used to identify the source of mistrustful activities are : Address,Total score and decision variable. The total score value has been calculated on the basis of multiple indvidual scores s1,s2,s3,s4. And the decision variable has been calculated on the basis of total score. The decision variable value is being taken as 0,1.and total score lies in between 0 to 22.

The address field consists of 7 cities : delhi,gurgaon,panipat,rohtak,jaipur,karnal,chandigarh

ssification Of Suspicious Cities On the basis on total weight Calculated

o.	Clusters	City	Total Weight	R
	Cluster 0	Delhi	168.0	1
	Cluster 1	Gurgaon	118.0	2
	Cluster 2	Chandigarh	116.0	3
	Cluster 3	Panipat	110.0	4
	Cluster 4	Karnal	58.0	5
	Cluster 5	Jaipur	29.0	6
	Cluster 6	Rohtak	29.0	7

Table 1 : Ranking according to severity of more suspicious activities

The table above shows the no of cities along with their data value interms of total weight of the transaction showing the no of suspicious activities or non suspicious activities . The rank provided to show that how much corresponding city is more questionable than the other cities. For insatance,the city delhi is having the highest value in terms of weight and hence provided the highest rank . So the largest no of mistrustful transactions in this city is the largest than the next city observed.Next Gurgaon is provided with the rank 2,which means having second largest no crimes occuring in this city.

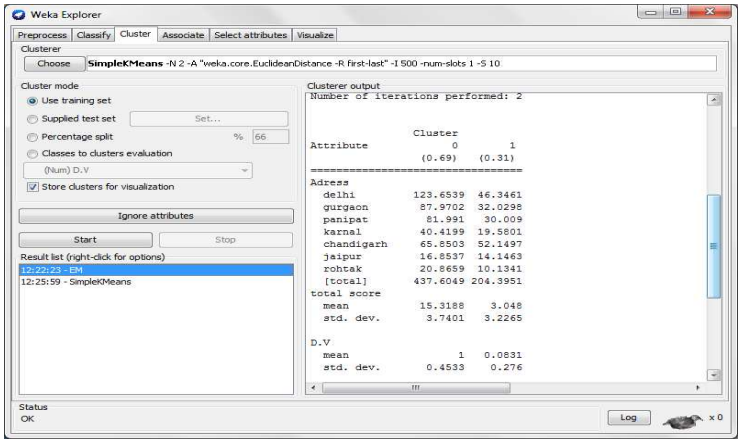

Figure 3 : No of Clusters formed and their evaluation

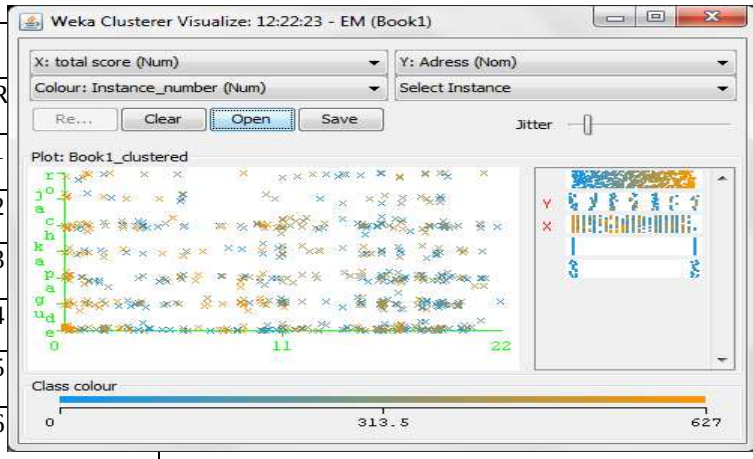

Figure 4 : Visualization of Clusters formed using EM

There are seven clusters in all which are shown in above figure; each cluster has been assigned its weight according to the total no of transactions in

each cluster. The largest the no of transaction in one cluster and if the transactions in it are greater in terms of their decision variable value then it is considered to be the most suspicious cities amongst all.The clusters are represented with the no of crosses.The DV value is represented with two distinct colors.The "Red" color span shows that the value is "1" which means activity is mistrustful. The other mark "Blue " color span shows thar the activity is trustful and no further observations are required.

5. Conclusion

The upshot received for the given problem helps in identification of chain of activities that contribute for occurrence of any kind of fiscal offence. No of cities which are prone to major bank fraud. The problem ultimately discovered the three major cities, out of analysed seven cities.Delhi, Gurgaon and Chandigarh having average weights 168.0, 118.0 and 116.0 respectively. These are the cities which are most vulnerable to financial crimes.Further considering Cluster0,Out of 168.0 transaction the total no of transactions, which are counterfeited, are 46 and rest of the activities are considered to be authentic. So for authentic transactions no further investigation is required. Similarly in cluster 1 which contains total of 116 activities,out of which the total no of transactions ,which are unfaithful, 32 and rest of the transactions are considred to be fair. So in cluster0 the total no of suspicious activities are much larger than that of second cluster which is cluster1. So cluster0 to which the city Delhi belongs is the most suspicious city in terms of financial offence.

REFERENCES

1. Financialfraud.pdf by John Howell & Co. Ltd.,August 2009
2. JERMY QUITTNER."AVOIDING CREDIT CARD FRAUD".
3. http://abcnews.go.com/business/_nancialSecurity/Story?id=89746andpage=12004
4. 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
5. M.R. Berthold et al, "Guide to Intelligent Data Analysis";
6. IJETAE_1112_112 (1).pdf,august 2011
7. Financial System Abuse, Financial Crime and Money Laundering Background Paper, February 12, 2001
8. file:///material/Thesis%20papers/Pragmatic%20Programming%20Techniques%20%20Fraud%20Detection%20Methods.htm
9. Vol_6(3)_311 - 322_Ogwueleka, FRANCISCA NONYELUM OGWUELEKA";
10. Trees, H.L.V. (2001). Detection, Estimation and Modulation Theory-Part I. John Wiley, New York.
11. 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.