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: <u>anu.dahiya71@gmail.com</u>, 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and loans, and thrift institutions to file a Currency Transaction Report (CTR)7 for any amounts that were deposited, withdrawn, transferred, or exchanged that exceeded \$10,000 in cash or coin (31 CFR 103.22).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.

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

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.

International States	ne 1 Calloni 18 Z	noest Page	A A	cornulas												0 - 11
aste 🥑	Calbri B Z	• 11 11 • 1 11 •	• A' A'	-												00 🕜 Co 🗞 s
ipboard is	BI	u - m -				ap Text	General					3- 3- 1	Σ AL	toSun * A	1	à 🍄 🍟
and the second se		Font	<u>≫-∆</u>		I 律律 函Me	rge & Center -	🧐 - % Nuri	• 58.43	Conditional Formatting	Forma as Table Shries	e * Styles *	Insert Delete For	at 2 cs	ear Filte Editing	& Fini	tà Spelling Purchase
A1		• (n	& Nam	e						-						
4	8	c	D	F	F G	н	1	1	ĸ		м	N O	P	0	E	- cut durand
Name	Are	Adress	Amount	no of days ch	eck bou PAN No	cibil score s	1 02		54	to	tal score	LV.		-	-	Catter Dented Oxfe
Reema da		25 delhi	45000	69	8 CGOUSLY	V 800	0	0	0	10	10	1				Main and Growt
striveta ra		25 sonipet	51000	16	11 QVGUW	G 900	1	0	0	10	11	1				
amily pan		26 panipat	52000	129	5 6208195	51 700	1	0	0	4	5	0				Do Nore
Joy		55 karnal	48000	14	11 AFRSLSW	1\ 900	0	0	0	10	10	1				Get Free Templates
purvi shui		60 chandigar	53000	181	8 8183DOH	N 500	1	4	0	10	15	1				Download Clip Art
shivani ka		35 chandigar	54000	20	2	0 900	1	0	1	2	4	0				is Get More
riya praka		22 sonipat	19000	40	3	0 800	0	0	1	2	3	0				Get Microsoft PowerPoint or Microsoft
anoop sor		28 panipat	50000	10	0	0 900	1	0	1	0	2	0				Outlook
0 diya kapu		29 delhi	49500	82	2	008 0	1	0	1	2	4	0				
1 siya khan		47 delhi	50000	101	12	0 700	1	0	1	10	12	1				
2 harsh chay		59 bhiwani	38000	187	16	0 500	0	4	1	10	15	1				
3 garima the		50 jalpur	31000	106	2 GLQSBRC	F 700	0	0	0	2	2	0			_	
4 Mukul der		19 chandigar	32000	56	4 PN30KF3	n 800	0	0	0	4	4	0				
5 harshit an		20 sonipat	56454	97	3 ZSM9UM	008 3	1	0	0	2	3	0				
ő amaira		19 karnal	56861	90	6 FU72N12	15 800	1	0	0	8	9	1			_	
7 Lai Chandi		52 rohtak	49747	76	1 OP063U	14 800	1	0	0	0	1	0			_	
a brijmohar		30 delhi	56421	166	9 6JG82H7	F! 600	1	0	0	10	11	1				-
9 Pandu Ma		33 delhi	50651	486	10 ML87LDT	P 20	1	9	0	10	20	1			_	Microsoft
J Jyotish Pa		31 delhi	50699	590	0 318QV4X	Yi -1	1	10	0	0	11	1			_	
1 Bhawna k		23 sonipat	51132	450	14 YJZY9640	T 20	1	9	0	10	20	1				Cffico
2 muskesh:		24 panipat	57869	355	6 J30VBLE	51 300	1	7	0	8	16	1			_	Da.Onice
3 muskan cł		25 panipat	49305	216	1 AFJIZNU	91 500	1	4	0	0	5	0			_	
a zoya akhri		28 delhi	57674	448	15	0 100	1	8	1	10	20	1			_	
5 monika sh		23 sonipat	55498	401	5	0 100	1	8	1	4	14	1			-	

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.

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

reprocess Classify Cluster Assoc	iate Select attribute	es Visualize					
Open file Open URL.	Open Di	Gene	rate	Undo	Edit	Save	
Filter							
Choose None						Apply	
Current relation Relation: Book1 Instances: 628	Sun	Attributes: 3 of weights: 628	Selected Name Missing	attribute : Adress : 0 (0%)	Distinct: 7	Type: Nominal Unique: 0 (0%)	
Attributes			No.	Label	Count	Weight	
			20.000	1 delhi	168	168.0	
All None	Invert	Pattern		2 gurgaon	118	118.0	
				3 panipat	110	110.0	
No. Name				4 karnal	58	58.0	
1 🔤 Adress			5 chandigarh		116	116.0	
2 total score				6 jaipur	29	29.0	
3 0.V			7 rohtak 29			29.0	
			Class: D.)	(Num)		 Visualize All 	
Res	nove]	168	118 110	58 58	29 29	
N- 6							

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,chandiga rh

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.

eprocess Classify Cluster Associate Select attrib	outes Visualize	
lusterer		
Choose SimpleKMeans -N 2 -A "weka.core.Eu	clideanDistance -R first-last" -I 500 -num-slots 1 -S 10	
lluster mode	Clusterer output	
 Use training set 	Number of iterations performed: 2	
Supplied test set Set		
Percentage split % 66	Cluster	
Classes to dustant evaluation	Attribute 0 1	
	(0.69) (0.31)	
(Num) D.V	Idrage	
Store dusters for visualization	delhi 123,6539 46,3461	1.0
	gurgaon 87.9702 32.0298	
Ignore attributes	panipat 81.991 30.009	
Charle Charles	karnal 40.4199 19.5801	
Stop	chandigarh 65.8503 52.1497	
esult list (right-dick for options)	jaipur 16.8537 14.1463	
2:22:23 - EM	rontak 20.8659 10.1341	
2:20:09 - Simpler means	total score	
	mean 15.3188 3.048	
	std. dev. 3.7401 3.2265	
	DW	
	mean 1 0.0831	
	std. dev. 0.4533 0.276	
	4 III III	

Figure 3 : No of Clusters formed and their evaluation

: C					Weka Clusterer Visualize: 12:22:23 - EM (Book1)			
ssinc	ation Of Suspicious	Cities On the basis on to	tal weight Calculated		X: total score (Num) Y: Adress (Nom)			
0.	Clusters	City	Total Weight	R	Colour: Instance_number (Num) Select Instance Re Clear Open Save Iitter			
	Cluster 0	Delhi	168.0	1	Plot: Book1_dustered			
	Cluster 1	Gurgaon	118.0	2				
	Cluster 2 Chandigarh 116.0		3					
	Cluster 3	Panipat	110.0	4				
	Cluster 4	Karnal	58.0	5	Class colour			
	Cluster 5	Jaipur	29.0	6	0 313.5 627			
	Cluster 6	Rohtak	29.0	7	Figure 4 : Visualization of Clusters formed using			
			1		EM			

Table 1 : Ranking according to severity of more suspicious activities

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 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

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 that 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/_nancialSecuri ty/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
- Financial System Abuse, Financial Crime and Money Laundering Background Paper, February 12, 2001
- file:///material/Thesis%20papers/Pragmatic%2 0Programming%20Techniques%20%20Fraud %20Detection%20Methods.htm
- 9. Vol_6(3)_311 322_Ogwueleka, FRANCISCA NONYELUM OGWUELEKA";
- Trees, H.L.V. (2001). Detection, Estimation and Modulation Theory-Part I. John Wiley, New York.
- 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.