

Tool for Detecting Irregular Patterns in the Use of Automated Teller Machine Card Using K-Means Algorithm

Benjamin Mensah, Dr. J B Hayfron Acquah, Sylvester Akpah

Abstract- The introduction of automated teller machine (ATM) cards in the financial sector has come with a lot of merits by making money almost available everywhere and at all times, The ATM cards are designed to mainly benefit people but when the ATM card gets into the hands of a wrong person with the secret PIN known by a popularly means called shoulder surfing or any other means the wrongful person at that time can perform a lot of unauthorized transactions. The task of this thesis is to come out with an algorithm to detect irregular ATM card usage patterns by means of questionnaires. The results from the analyses will aid come up with an algorithm to detect irregular patterns from transactions carried out by ATM card users. The algorithm to be developed will keep in memory of all withdrawal patterns deemed to be regular and irregular, the algorithm will decide from the definition of regular transactions to approve or block current transactions going on in an ATM. Transactions deemed irregular by the algorithm will prevent the current transactions and prompt or advise the customer performing that transaction to visit the banking hall for further transactions, this action by the algorithm will prevent or reduce the rate of ATM card fraud by preventing transactions deemed suspicious and at the same time giving advice to good customers performing unusual transaction to visit the banking hall. The aim of the thesis is to help minimize the harm that can be done with an ATM card within twenty four hours.

KEYWORDS: Clustering techniques and algorithm, anomaly techniques and K-means clustering.

1.0 Introduction

Automated Teller Machine (ATM) is an electronic banking outlet, which allows clients to complete basic transactions without the aid of a branch representative or teller.

Two main types of automated teller machines are currently in the system. They are the basic and complex type. The basic units allow the client to only withdraw cash and receive a report of the account's balance. The more complex machines will accept deposits, facilitate credit card payments and report account information. To access the advanced features of the complex units, one will usually need to be a member of the bank that operates the machine.

An ATM card sometimes called debit card, customer card, main card, cash card or debit card are cards that are specially developed by banks and given to customers mostly on request to enhance clients access the automated teller machine for the purposes of transacting business such as cash withdrawals, mini statement, and for some advanced ATM's for cash deposit. Many other types of simple banking activities can be carried out by ATM too. The payment card are normally cards with the features enabled, and are usually a credit, debit, a limited-use ATM or other forms of cards. The same banks at different locations when networked will allow for intra ATM transactions to take place. Different banks can also permit each other for interbank ATM transactions.

ATM cards can perform activities of some already existing means of financial transaction such as merchants' card workstations that performs ATM functions not allowing cash transactions mostly cash drawer and it is mostly termed as mini ATMs. The workstations can also perform the task as Cashless scrip ATMs by cashing the fund transfer receipt at the merchant's Cashier.

The ATM cards have come along with numerous merits to the financial institutions and their clients. The functions of ATM cards involve algorithms that enable end users to perform functions in definite manner. When one default the rules in the use of the ATM card, the transaction to be performed is prevented. These algorithms behind the ATM card usage cannot be modified by the card bearers but the user of the card can only perform particular defined operations at the ATM machines.

As the evolution of ATM cards have brought along numerous merits it also have some demerits associated with it. When ones ATM card is stolen or gets to the wrong hand with the secret PIN known, the wrongful owner at that time can perform transactions with the card that may be disastrous to the right owner.

This research is therefore aimed at improving on existing algorithms for tracking irregular patterns for the use of ATM cards by clustering technique.

By the use of the tool to be developed, most of the transaction patterns of ATM cards will be known and distinction drawn between what will be deem as regular and what will be deemed as irregular. Must people will know of their missing ATM cards within twenty four hours, the thesis therefore seeks to address or minimize the risk that can happened within twenty four hours with an ATM card in a wrongful hand

1.1 Problem Statement

The introduction of ATM cards into the banking system has come to reduce tremendously the work done at the banking halls, now fewer queues are seen at most banking halls

because ATM cards can perform most of the usual functions done at the banking halls, nevertheless apart from it numerous benefits comes along with some disadvantages if the ATM cards gets to the wrong hands with the secret pin known. The wrongful owner at that time can perform all transactions with the card that may be dangerous to the rightful owner.

The thesis aims at tracking some abnormal transaction patterns in the use of ATM cards. The information from the ATM will assist in gathering regular transaction of ATM cards and also monitor unusual transactions that fall out the regular transactions, financial institution can use the outcome of the thesis to improve upon the current algorithm being used by the ATM.

The algorithm will go a long way to reduce ATM card fraud currently going on in the system and also improving the trust in ATM system.

1.2 Research Questions

The main research question that will serve to guide this thesis is "what are the most irregular withdrawal patterns of automated teller machine cards users?"

1. What are automated teller machines and cards?
2. What withdrawal patterns are deemed irregular?
3. How do one know irregular patterns?
4. What algorithm can best be used to detect irregular patterns?

1.3 Objective of the Study

The main objective of the study is to come up with a reliable algorithm for detecting irregular patterns in the use of ATM cards by using questionnaires to ascertain the needed information to help come out with the algorithm.

1.4 Literature Review

Anomaly Detection: is the process of identifying irregular patterns of situations. It mainly deals with comparing day to

day transactions data stored to current or new data pattern. When the comparison between the two set of information is done and any deviation is tracked, that new deviation is called anomaly detection.

K-means clustering: Is an unsupervised learning algorithm used to trace or predict well known clustering problems. The process a non-difficult method of classifying a given data set through a certain number of clusters (assume k clusters) fixed a priori)

According to A. M. Riad et al (2013) of Mansoura University, Egypt, they presented a thesis based on the fact that with the ever increasing amount of new attacks happening currently in the world will keep increasing because of the base-rate fallacy of the amount of false alarms that keep increasing. Another issue with detection is the slow rate of which crimes are detected. Of late, most crimes are usually known after it had been executed, this makes vital information very unsecured to wrongful people.

The K-means by using Cluster 3.0 tool was applied to tackle the task. The steps that follows were used: false alarm rate and detection rate. The detectable number divided by the sum total number of attacks is referred as detection rate. The sum total of 'normal' patterns classify occurrences divided by the total number of 'normal' patterns is called false alarm rate.

The sum total of malicious properly classified as malicious: True Positives (TP); the sum total of benign programs properly classified as benign is termed as: True Negatives (TN); the sum total of benign programs incorrectly classified as malicious is called: False Positives (FP); the number of malicious falsely classified as benign is called: False Negative (FN)

$$\text{Detection Rate (DTR)} = \frac{TP}{(TP + TN)} \text{ and False Alarm Rate (FPR)} = \frac{FP}{(TN + FP)}$$

Can be explained as follows:

The entire amount of normal patterns: 60593, the entire amount of 'attacks' patterns: 250436. The entire amount of all detection: 311029.

$$\text{Detection Rate (DTR)} = \frac{TP}{(TP + TN)}$$

$$\text{False Alarm Rate (FPR)} = \frac{FP}{(TN + FP)}$$

| Clustering Technique | | DOS | Probe | U2R | R2L | |
|----------------------|----------------|--------|--------|---------|--------|--|
| K-Means K=4 | Detection rate | 0.9993 | 0.0656 | 0.00005 | 0.064 | |
| | False Alarm | 0.001 | 0.004 | 0.00007 | 0.0001 | |
| K-Means K=5 | Detection rate | 0.9766 | 0.0659 | 0.00061 | 0.381 | |
| | False Alarm | 0.003 | 0.013 | 0.0004 | 0.0022 | |
| K-Means K=6 | Detection rate | 0.9643 | 0.0654 | 0.00099 | 0.4997 | |
| | False Alarm | 0.004 | 0.026 | 0.008 | 0.001 | |

From table 2.1, the sum total of regular pattern is 60593 and that of irregular patterns is 250436 with the grand total being 311029.

Table 2.1: Experiment Results of K-Means

The research outcome indicate that K-means when k=4 gives the optimum prediction rate on the high and false tare is less than others.

Figure.2.1 demonstrates the rate of detection and the wrong alarm for the four criteria of attacks (DOS, Probe, R2L, U2R) with different clusters (k=4, 5, 6)

The outcome of the experiment showed that K-means when $k=4$ is the preeminent as finding rate is high and false alarm rate is less than others.

Figure 2.1 also illustrates the detection rate and false alarm for the four categories of attacks (DOS, Probe, R2L, U2R) with different clusters ($k=4, 5, 6$)

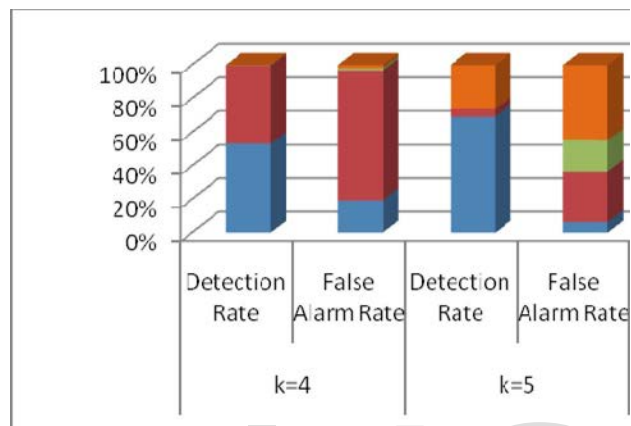


Figure 2.1: Detection rate and false alarm

1.5 Methodology

The research is categorized into the following sub sections:

- Data collection introduction
- Data processing and analysis
- Delineation of study
- Preliminary definition of terms

1.5.0 Data Collection

A total of hundred and thirty questionnaires were designed and administered. Out of the total number of questionnaires administered, a sub total of hundred were retrieved. The remaining thirty were not retrievable because as at the time of the analysis the respondents have not returned them.

Criteria were drawn to select respondents for the questionnaire which made the mode of selection random and

selective sampling. The following were the laid down criteria.

- The respondent should be an ATM card user. This is because only ATM card users are needed for the research
- The respondent should at least have used ATM card for not less than six month. This condition enables respondent to have a clear ATM card usage pattern
- The respondent should have a clear or enough memory of their ATM card usage pattern and
- Literate were preferred because the questionnaire was given to the respondent and later taken back, meaning issuance of the questionnaire will not be available for assistant/ interpretation.

Two main statistical techniques were used for choosing respondents. They were random sampling and bias/selective sampling. The random sampling was used to give equal chance to all legible respondents and the bias sampling was used because not all customers of banks uses ATM cards and also only customers who had used ATM card for more than six months were legible for the questionnaire so that they (customers) might have a clear and enough memory of their withdrawals patterns. Also bias sampling was used because illiterate were not selected as respondents because they could most of the time not use the ATM cards.

These conditions and the selection were achieved by verbally making polite enquiry before issuing questionnaire to the respondent. By these criteria most willing respondents were not issue questionnaire.

1.5.1 Data Processing and Analysis

Group each question in questionnaire according to the question number, after the grouping by question numbers make a sub group of each question by the answers retrieved from each question number according to similarity.

IBM SPSS and Microsoft Excel were used for the data processing and analyses. Code the outcome of the pattern of how questions are answered in the questionnaire in the variable view of the IBM SPSS, the k-means clustering tool

in the IBM SPSS is used for the numerical analysis and Microsoft Excel is used for the graphical interpretation. MATLAB is used to show the clusters of the various questions patterns and also Microsoft Excel is used for the graphical interpretation of the data.

The qualitative answers from the questionnaire are grouped and coded into numbers since the tool can only perform analysis on numbers.

Group data from each question are well labelled in each column on the spread sheet. Pie chart is developed from the data gathered from each question.

1.5.2 Delineation of Study

The IBM SPSS was used for the pattern recognition by a tool in the software called k-means cluster. Before implementing the k-means cluster, all needed coding in the variable view should be done and all data entry in the data view should have been done.

1.6 Summary of Major Finding

The following are the major outcomes of the research:

- The time interval in a day most withdrawals are made by ATM cards holders is between 6:00am-6:59pm. There were insignificant withdrawals made outside this time interval.
- Number of times withdrawals of cash are usually made within an hour from the analyses is 1-2 times in an hour.
- Number of most withdrawals within a day from the analyses is 1-3 times within a day.
- How often withdrawals are made to the maximum limit allowed. As to if people withdraw up to the maximum or not, a clear pattern was not known because people had different amount of withdrawal patterns though from the questionnaire more than half withdraw up the total amount permitted by the individual banks, there were also a considerable number of people who had no particular amount they withdraw at any time.

- Number of different ATMs used within an hour by most people for transaction from the analyses is one.
- Number of different ATMs used within a day for transaction by most people from the analyses is one or at most two.
- Radius within which ATMs transaction are carried out within a day by most people from the analyses is within 0.01km to 4.00 km

1.7 Proposed algorithm

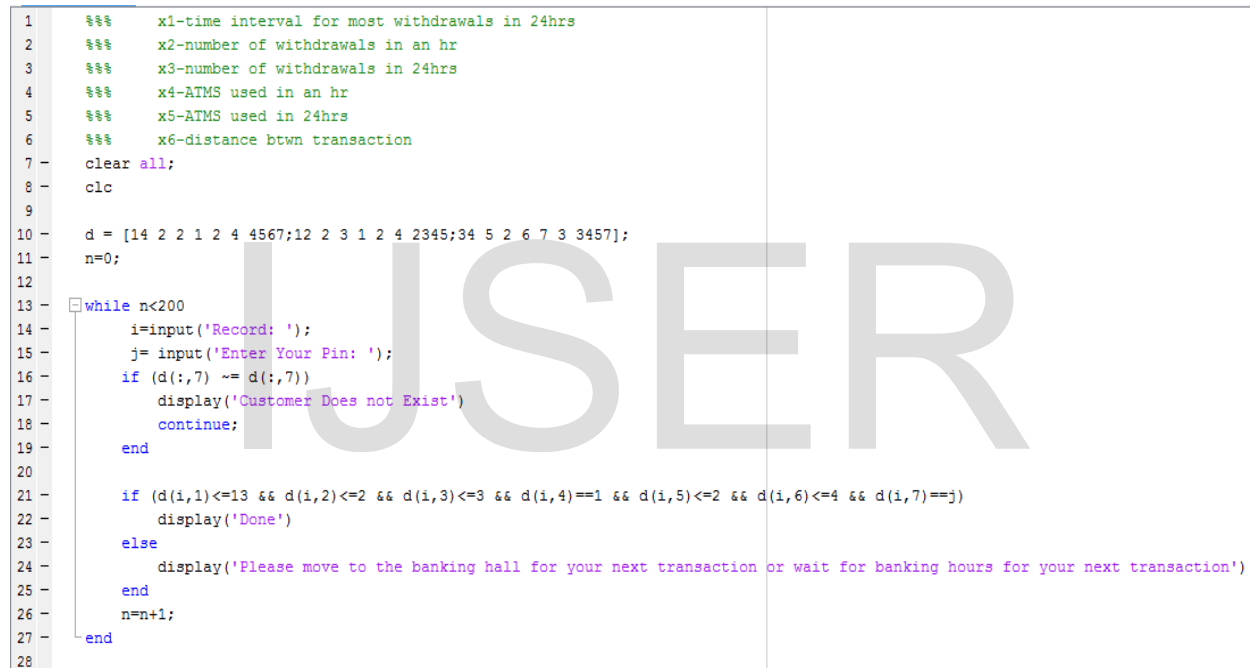
The proposed algorithm for detecting irregular patterns in the use of ATM cards is as follows:

1. Read the given data
2. Re-categorize the data into six groups as time interval for most withdrawals within a day, number of times withdrawals are made within an hour, number of times withdrawals are made within a day, different ATM used for transactions within an hour, different ATM used for transaction within a day, radius in which ATM transactions are most performed
3. For group one, if time for withdrawal is from 6:00am to 6:59pm then allow transaction else move to the last step
4. For group two, if withdrawal within an hour is two or less then allow withdrawal else move to the last step
5. For group three, if withdrawal within a day is three or less then allow withdrawal else move to the last step
6. For group four, if number of ATMs used for transaction within an hour is one allow else move to the last step
7. For group five, if number of ATMs used within a day is two allow transaction else move to the last step
8. For group six, if radius of transaction between ATMs is 0.01km to 4.00km and for a time interval of 24 hours allow transaction else move to the last step
9. Don't allow transaction (please move to the banking hall for your next transaction or wait for banking hours for your next transaction)

1.8 Simulation of Algorithm

The algorithm designed is stimulated by MATLAB. Figure 1.0 shows the algorithm implemented from MATLAB R2013. The algorithm implemented in MATLAB takes information in the form of a matrix. The matrix formed are row matrix with each column in the matrix depicting different condition in the algorithm. The following are the interpretation of each column in the matrix.

- the first column value stands for “time interval for withdrawals within 24 hours”
- the second columns value stands for “number of most withdrawals in an hour”



```

1  %%% x1-time interval for most withdrawals in 24hrs
2  %%% x2-number of withdrawals in an hr
3  %%% x3-number of withdrawals in 24hrs
4  %%% x4-ATMS used in an hr
5  %%% x5-ATMS used in 24hrs
6  %%% x6-distance btwn transaction
7  clear all;
8  clc
9
10 d = [14 2 2 1 2 4 4567;12 2 3 1 2 4 2345;34 5 2 6 7 3 3457];
11 n=0;
12
13 while n<200
14     i=input('Record: ');
15     j= input('Enter Your Pin: ');
16     if (d(:,7) ~= d(:,7))
17         display('Customer Does not Exist')
18         continue;
19     end
20
21     if (d(i,1)<=13 && d(i,2)<=2 && d(i,3)<=3 && d(i,4)==1 && d(i,5)<=2 && d(i,6)<=4 && d(i,7)==j)
22         display('Done')
23     else
24         display('Please move to the banking hall for your next transaction or wait for banking hours for your next transaction')
25     end
26     n=n+1;
27 end
28

```

- the third column value stands for “number of withdrawals within 24hours”
- the fourth column value stands for “number of ATMs used in an hour”
- the fifth column value stands for “number of ATMs used within 24hours”
- the sixth column value stands for “distance between different ATM used for transactions within a day.
- The seventh column value stands for “ the PIN to be entered”

When all seven requirements are satisfied, the customer is allowed to carry on with his/hers transaction, when even one condition is not fulfilled the entire transaction is prevented and the customer is advised to enter the banking hall for the appropriate transactions or wait till the official banking hours.

Figure 1.0: illustrate the editor view of the algorithm implemented in MATLAB

Figure 1.1 illustrate the number of times one usually withdraw cash within an hour, from figure 1.1, there are two main clusters that were found from the data analyzed, thus

cluster one (1) and cluster two (2), it can be seen from figure 1.1 that cluster one has maximum of data in that group with cluster two being the outlier of having extreme minority of data members.

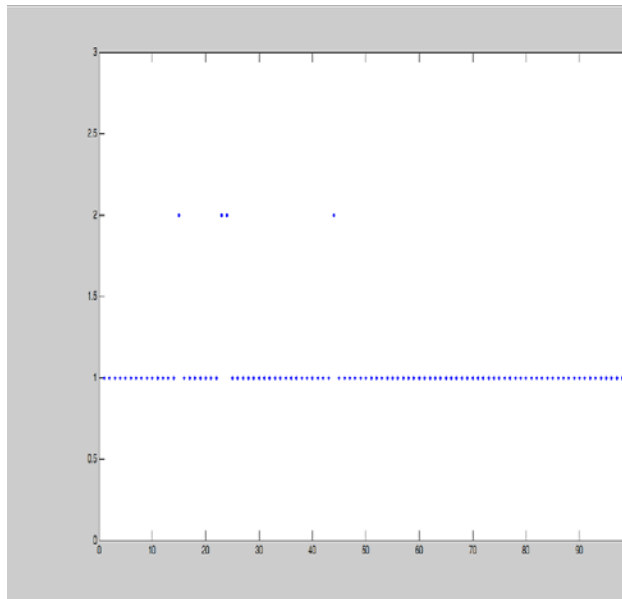


Figure 1.1: illustrate data patterns from the questionnaire of time most withdrawals are made in the day

Legend

1= time from 6:00am-6:59pm

2= time from 7:00pm-5:59am

Figure 1.2: illustrate the data pattern from the questionnaire of most number of withdrawals made in an hour

Legend

1 = 1-2 withdrawals within an hour

2= 3-4 withdrawals within an hour

3= 4 and above withdrawals within an hour

- Number of most withdrawals within a day from the analyses is between 1-3 times

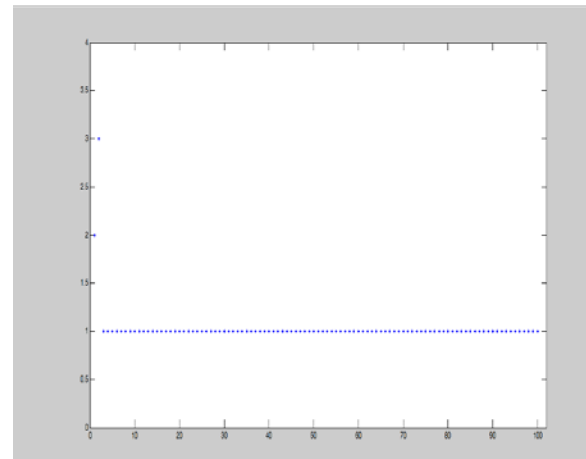


Table 1.2: illustrate the initial Cluster Centers of number of times withdrawals are made within a day

1.9 Limitations

There were no actual data from the financial institutions because of the confidentiality they hold for their clients. This was a big blow to the research since such data could not be altered by the respondents as compared to the questionnaire designed and administered to the respondents.

COMPETING INTERESTS:

Authors have confidently declared the non-existence of any competing interests.

References

- 1Faculty of Computer Science and Information Systems, Mansoura University, Egypt
- 2Faculty of Computer Science and Information Systems, Zagazig University, Egypt
- 3Faculty of Engineering Mansoura University, Egypt, Kanishka Bhaduri², Kamalika Das³, and Katharina Morik¹ of 1 TU Dortmund, Computer Science, LS 8, 44221 Dortmund, Germany2 Netflix Inc., Los Gatos, CA 94032, USA3 UARC, NASA Ames, CA 94035, USAShen, R. Tong, and Y. Deng, "Application of classification models on credit card fraud detection," June 2007. International Journal of Computer Networks & Communications (IJCNC) Vol.5, No.5, September 2013

- Shen, R. Tong, and Y. Deng, "Application of classification models on credit card fraud detection," June 2007
- Abhinav Srivastava, Amlan Kundu, Shamik Sural, Arun K. Majumdar, "Credit Card Fraud Detection using Hidden Markov Model," IEEE Transactions On Dependable And Secure Computing, vol. 5, Issue no. 1, pp.3748, January-March 2008.
- Abhinav Srivastava, Amlan Kundu, Shamik Sural, Arun K. Majumdar, "Credit Card Fraud Detection using Hidden Markov Model," IEEE Transactions On Dependable And Secure Computing, vol. 5, Issue no. 1, pp.3748, January-March 2008. International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 2, February 2014
- Aihua Shen, Rencheng Tong, Yaochen Deng, Application of Classification Models on Credit Card Fraud Detection, 2007 IEEE.
- Amlan Kundu, Suvasini Panigrahi, Shamik Sural and Arun K. Majumdar, "Credit card fraud detection: A fusion approach using Dempster-Shafer theory and Bayesian learning," Special Issue on Information Fusion in Computer Security, Vol. 10, Issue no 4, pp.354- 363, October 2009.
- Angiulli, F., Basta, S., Lodi, S., Sartori, C.: A Distributed Approach to Detect Outliers in Very Large Data Sets. In: Proc. of Euro-Par'10. pp. 329–340 (2010)
- Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM Comp. Surveys 41(3), 1–58 (2009)
- Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM Comp. Surveys 41(3), 1–58 (2009)
- Das, K., Bhaduri, K., Votava, P.: Distributed anomaly detection using 1-class SVM for vertically partitioned data. Stat. Anal. Data Min. 4(4), 393–406 (2011)
- Das, S., Matthews, B., Srivastava, A., Oza, N.: Multiple kernel learning for heterogeneous anomaly detection: algorithm and aviation safety case study. In: Proc. of KDD'10. pp. 47–56 (2010)
- Hodge, V., Austin, J.: A survey of outlier detection methodologies. A. I. Review 22(2), 85–126 (2004)
- Hung, E., Cheung, D.: Parallel Mining of Outliers in Large Database. Distrib. Parallel Databases 12, 5–26 (2002)
- J.J. Peersman "Preventing Data Breaches by Proactive Data mining" of Utrecht University, December 28, 2012
- Jon T. S. Quah and M. Sriganesh, Real Time Credit Card Fraud Detection using Computational Intelligence, Proceedings of International Joint Conference on Neural Networks, Orlando, Florida, USA, August 2007. Kou, Y.
- Lu, C.-T., Sirwongwattana, S., Huang, Y.-P.: Survey of fraud detection techniques. In: Proceedings of the 2004 IEEE International Conference on Networking, Sensing and Control, Taipei, Taiwan (2004).
- K. Hanumantha Rao¹, G. Srinivas², Ankam Damodhar³ and M. Vikas Krishna⁴ ^{1,2,3,4}Sri Indu College of Engineering and Technology, Hyderabad, India Khyati Chaudhary¹ Bhawna Mallick² of ¹, ²Galgotias College of Engg. & Technology, Greater Noida International Journal of Computational Engineering Research. / ISSN: 2250–3005. International Journal of Computer Science and Telecommunications [Volume 2, Issue 3, June 2011]

- Lozano, E., Acuna, E.: Parallel algorithms for distance-based and density-based outliers. In: ICDM'05. pp. 729–732 (2005)
- Raghavendra Patidar, Lokesh Sharma, “Credit Card Fraud Detection
- Tej Paul Bhatla, Vikram Prabhu & Amit Dua “Understanding Credit Card Frauds,” 2003.
- Using Neural Network”, International Journal of Soft Computing and Engineering (IJSCE) June 2011.
- Y. Sahin, E. Duman “Detecting Credit Card Fraud by ANN and Logistic Regression” 2011.
- Y. Sahin, E. Duman, Detecting Credit Card Fraud by ANN and Logistic Regression, ©2011 IEEE

IJSER