

REALTIME FRAUD DETECTION IN THE BANKING SECTOR USING DATA MINING TECHNIQUES/ALGORITHM

S. N. John

Department of Electrical and Information
Engineering
Covenant University,
Ota, Ogun State, Nigeria.
samuel.john@covenantuniversity.edu.ng

C. Anele

Department of Information Technology
National Open University of Nigeria.
Victoria Island, Lagos, Nigeria
ahneyile0007@gmail.com

Okokpuije Kennedy O.

Department of Electrical and Information
Engineering
Covenant University,
Ota, Ogun State, Nigeria.
kennedy.okokpuije@covenantuniversity.edu.ng

F. Olajide

Department of Computer and Information
Sciences
Covenant University,
Ota, Ogun State, Nigeria.
funminiyi.olajide@cu.edu.ng

Chinyere Grace Kennedy

Department of Computer Science and
Engineering,
Ewha Womans University,
Seoul, South Korea.
gkennedy@ewhain.net

Abstract—The banking sector is a very important sector in our present day generation where almost every human has to deal with the bank either physically or online. In dealing with the banks, the customers and the banks face the chances of been trapped by fraudsters. Examples of fraud include insurance fraud, credit card fraud, accounting fraud, etc. Detection of fraudulent activity is thus critical to control these costs. This paper hereby addresses bank fraud detection via the use of data-mining techniques; association, clustering, forecasting, and classification to analyze the customer data in order to identify the patterns that can lead to frauds. Upon identification of the patterns, adding a higher level of verification/authentication to banking processes can be added
Keywords: Data mining techniques, banking sector, fraud, and authentication.

Key words: Data Mining, Authentication, Real Time Fraud, Banking Sector.

I. INTRODUCTION

According to The American Heritage dictionary, second college edition, fraud is defined as a deception deliberately practiced in order to secure unfair unlawful gain. Fraud detection is the recognition of symptoms of fraud where no prior suspicion or tendency to fraud exists. Examples include insurance fraud, credit card fraud and accounting fraud.

Data from the Nigeria Inter-Bank Settlement System (NIBSS) has revealed that fraudulent transactions in the banking sector at its peak [1]. Fraud has evolved from being committed by casual fraudsters to being committed by organized crime and fraud rings that use sophisticated methods to take over control of accounts and commit fraud. Some 6.8 million Americans were victimized by card fraud in 2007, according to Javelin research [2]. Such fraud on existing accounts accounted for more than \$3 billion in losses in 2007. The Nilson Report estimates the cost to the industry to be \$4.84 billion. [3] Javelin estimates the losses at more than six times that amount – some \$30.6 billion in 2007. [2] Of course, fraud is not a domestic product as it's everywhere. For instance, card fraud losses cost UK economy GBP 423 million in 2006. Credit card fraud accounts for the biggest cut of the \$600 million that airlines lose each year globally.

Card losses top ZAR 50 million a year in South Africa (US\$6.3 million), according to the South African Card Fraud Forum. The good news is that the numbers tend to be slightly down from previous years, especially in the US. But the bad news on the other hand happens to be that hackers, identity thieves and money launderers are fighting back by focusing on different channels and keeps coming up with new types of attacks that traditional fraud management strategies were not designed to address.

A. Data Mining Model

The descriptive model points out the patterns or relationship in data and goes ahead to explore the properties of the data examined. Figure 1 shows the data mining models and tasks.

B. Data Mining Tasks

The use of result from data mining goes a long way to determine the data mining task to be performed. Data mining tasks are categorized as follows [4].

1. Exploratory Data Analysis: it is as simple as the name implies. Data is explored without any clear idea of what is being looked for.
2. Descriptive Modeling: It describes all the data involved and models the relationship between every single variable.
3. Predictive Modeling: This model permits the value of one variable to be predicted from the known value of other variables.

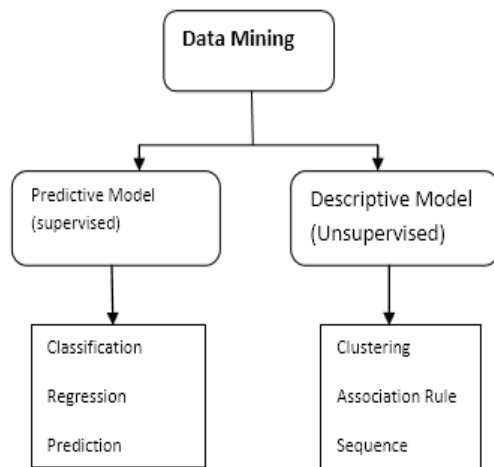


Figure 1 Data Mining Models and Tasks

4. Discovering Patterns and Rules: This is concerned with the detection of patterns and unraveling of fraudulent behavior by exposing regions where data points significantly different from the rest.
5. Retrieval by Content: Here, patterns which are similar to the pattern of interest in the data set are figured out. It is mostly applicable to text and image data sets.

II. LITERATURE REVIEW

Banks are experiencing challenges in protecting the online/internet banking channel. The challenge is in keeping customer's account secure while avoiding complexity in the login process [5]. However the myriad of passwords, hardware token devices and other out-of-bound communication tools introduced by some banks has greatly discouraged some customers. Well it is obvious that security focuses less on customer convenience, but there can still be an improvement.

Fractal makes use of sophisticated analytical models and rules to correctly identify suspicious behavior. But customers can be greatly frustrated if a transaction is incorrectly declined-meaning that banks need to select fraud prevention systems that deliver the lowest levels of false positives.

Online Banking Authentication is getting more complex as new threats are discovered and the technology needs to secure users against them. Authentication used to be a simple password but because of growing threats over the years it has grown from that to password with numbers, then to password with numbers, symbols and special character, to knowledge based questions and finally to the current state of external devices (token) and communication channels to verify transactions.

The solution proposed therefore is to merge the security strength of an authentication server with the logic and accuracy of a fraud detection system to identify suspect behavior and step up authentication, or on the other hand where normal behavior is recognized, to step down authentication so that the customer is not unnecessarily inconvenienced – thus achieving low risk and high customer satisfaction. Fractals can help institutions decide:

- What should trigger increased or decreased authentication?
- Should access be allowed when high risk authentication is not available, or should restrictions be implemented?
- What constitutes a “normal” profile for a customer?

In their conclusion, it was said that using fractals to enable Intelligent Risk Authentication means these strategies can be crafted, implemented and managed quickly, easily and with the end goal of greater customer satisfaction without greater risk.

A. Overcoming Losses to Fraud

A fraud survey conducted in 2011 by FICO states that the following is required for fraud losses to be reduced drastically [6]

1. Fraud detection in real time.
2. Analytics
3. Workflow
4. Efficient rules engine

III. USING DATA MINING TECHNIQUES IN REAL TIME FRAUD DETECTION

In our proposed methodology, we supposed that a fraud detecting system has the following objectives:

- To eliminate real time fraud to the lowest level.
- To increase the confidence of customers in the banking system especially for online transactions.
- To discourage fraudsters (both present and intending ones)

There are several data mining techniques, and most have been used in data mining research projects. Amongst these developed techniques include classification, clustering, association, prediction, and sequential patterns.

A. Classification

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Fraud detection and credit risk applications are particularly well suited to this type of analysis. The data classification process involves learning and classification.

B. Clustering

Clustering can be said to be the identification of similar classes of objects. In this technique, transactions with similar behavior are combined into one group. Clustering can be used as preprocessing approach for attribute subset selection and classification [7]. For instance: The customer of a given geographic location and of a particular job profile demand a particular set of services, like in banking sector the customers from the service class always demand for the policy which ensures more security as they are not intending to take risks, likewise the same set of service class people in rural areas have the preferences for some particular brands which may differ from their counterparts in urban areas. This information will help the organization in cross-selling their products, instead of mass pitching a certain “hot” product, the bank's customer service representatives can be equipped with customer profiles enriched by data mining that help them to identify which product and services are most relevant to callers.

This technique will help the management in finding the solution of 80/20 principle of marketing, which says: 20% of your customers will provide you with 80% of your profits, then the problem is to identify those 20% and the techniques of clustering will help in achieving the same.

C. Association Rule

The central task of association rule mining is to find sets of binary variables that co-occur together frequently in a transaction database, while the goal of feature selection problem is to identify groups that are strongly correlated with each other with a specific target variable.

Association rule has the several algorithms like: APRIORI, CDA, and DDA. Association rules are if/then statements that help uncover relationships between seemingly unrelated data in a relational database or other information repository. An example of an association rule can be 'if a customer buys a dozen eggs, he is 80% likely to also purchase milk.' There are two parts in an association rule, an antecedent (if) and a consequent (then).

D. Prediction

The prediction as its name implies is one of the data mining techniques that discover relationship between independent variables and relationship between dependent variables. For instance, prediction analysis technique can be implemented in the banking sector to predict fraud. Money can be seen as the independent variable while the individual (fraudster) could be seen as the dependent variable. Then based on historical data, we can draw a fitted regression curve that is used for attempted fraud prediction. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. In data mining, independent variables are attributes already known and response variables are what we want to predict. Unfortunately, many real-world problems are not simply predictable. Types of Regression Techniques are as follows:

- Linear Regression
- Multivariate Linear Regression
- Nonlinear Regression

Multivariate Nonlinear Regression

E. Sequential Patterns

Sequential patterns analysis is one of data mining techniques that seek to discover similar patterns in data transaction over a business period. The uncovered patterns are used for further business analysis to recognize relationships among data.

IV. SECURED BANKING TRANSACTIONS

The bank is an organization which provides facilities for acceptance of deposits, and provision of loans. In the ancient days, wealth was usually deposited in temples and treasuries. The earliest banks were used exclusively by rulers to fund the more important and larger festivals and for building expenses.

Now with a paradigm shift that made banking available to all, there is no doubt that some folks who do not have funds in the bank would look for means to obtain what is not theirs, hence the need for adequate secured banking procedure.

The fraud detection design model is an implementation of the reviews of all the above discussed data mining techniques. The design also put into consideration the following facts:

- Fraudsters have taken their time to study the operational procedures of banks.
- Some of them are so close to their potential victims that they can forge their signatures without doubt.

- The possibility of the existence of a lookalike also puts some withdrawal processes to question.

Since there has been an existing record of each customer's transaction history with the bank, that forms our data warehouse as shown in Figure 2.

N/B: Association rules are created by analyzing data for frequent if/then patterns and using the criteria support and confidence to identify the most important relationships. The **support** is an indication of how frequently the data items appear in the database. The **confidence** indicates the number of times the if/then statements have been found to be true.



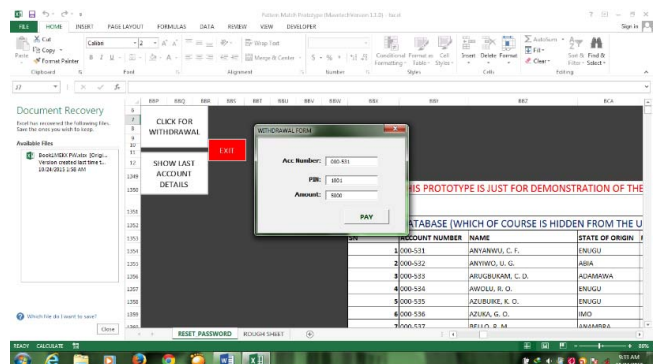
ACCOUNT NUMBER	NAME	STATE OF ORIGIN	PIN	SECRET QUESTION	VALUE	BALANCE
1000-531	ANYANWU, C. F.	ENUGU	1001	ELDEST NIECE	CHINESE	16,600.00
2000-532	ANYIBO, U. G.	ABIA	1002	DADY'S BIRTH MONTH	NABON	87,000.00
3000-533	AKURURUANA, C. D.	AKAMARA	1003	BEST COLOUR	GREEN	120,000.00
4000-534	AKUOLU, R. O.	ENUGU	1004	BEST SUBJECT	ENGLISH	580,000.00
5000-535	AZUBUIKE, K. O.	ENUGU	1005	BEST COLOUR	YELLOW	4,699,900.00
6000-536	AZUKA, G. O.	IMO	1006	DADY'S BIRTH MONTH	JUNE	16,600.00
7000-537	BELLO, R. M.	ANAMBRA	1007	DADY'S BIRTH MONTH	MAY	87,000.00
8000-538	CHARLTON, P. G.	ANIGBO	1008	BEST COLOUR	BLUE	120,000.00
9000-539	CHIEKE, V. N.	KEF	1009	ELDEST NIECE	NABON	16,600.00
10000-540	HARTINA, L. I.	PATATA	1010	ELDEST NIECE	JOY	87,000.00
11000-541	CHINWE, B. C.	ENUGU	1011	ELDEST NIECE	NAKKE	6,120,000.00
12000-542	CHINWE, I. B.	ENUGU	1012	ELDEST NIECE	KEVIN	580,000.00
13000-543	CHINWE, G.	HARTINA	1013	ELDEST NIECE	KEVIN	3,699,900.00

Figure 2 Shows account holders with their respective account balance

On studying the data listed in Figure 2 above, observations were made in detecting data items that are associated or correlated with each other and which was not obvious previously. For instance, if it was noticed in the data warehouse that Mr Anyanwu (no 1 customer on the data-list) never exceeds a particular range of amount (N5,000.00) every time he goes to the bank to withdraw. This is also our **Support** in relation to association rule.

Secondly, our **Confidence** with respect to association rule is the fact that the withdrawal of N5,000.00 occurred more than half of the times that Mr. Anyanwu withdrew cash from the bank.

Concluding from the two premise made above which also agrees with association rule, we can then say that it is rarely possible for the customer in question to withdraw above N5000 and it won't be absurd if "alarm" is raised whenever the database records that MR Anyanwu wants to withdraw above the accustomed amount and then prompt a pin request or any other authentication as designed by the IT team. And same applies to other customers of the bank who likewise maintains a particular range of accustomed amount.



ACCOUNT NUMBER	NAME	STATE OF ORIGIN
1000-531	ANYANWU, C. F.	ENUGU
2000-532	ANYIBO, U. G.	ABIA
3000-533	AKURURUANA, C. D.	AKAMARA
4000-534	AKUOLU, R. O.	ENUGU
5000-535	AZUBUIKE, K. O.	ENUGU
6000-536	AZUKA, G. O.	IMO
7000-537	BELLO, R. M.	ANAMBRA

Figure 3 Photo of account debiting screen

The Figure 3 above displays the screen when the teller [cashier] enters the figure [N5,000] to debit and pay Mr Anyanwu with acct no 000-531. On entering the above figure [N5,000] and clicking ‘PAY’, which is within the accustomed threshold the transaction is consummated with ease see Figure 4.

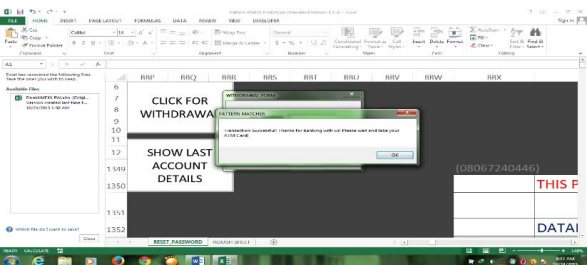


Figure 4 shows a successfully consummated transaction

But in a situation whereby an imposter who resembles our No.1 customer and probably knows his signature and knows Mr. Anyanwu’s account number and ”secrete PIN” attempts to withdraw from his account, the imposter would not be aware of the transaction limit that his victim is acclimatized to, and in so doing he/she would definitely overshoot. The Figure 5 indicates, debit above threshold limit.

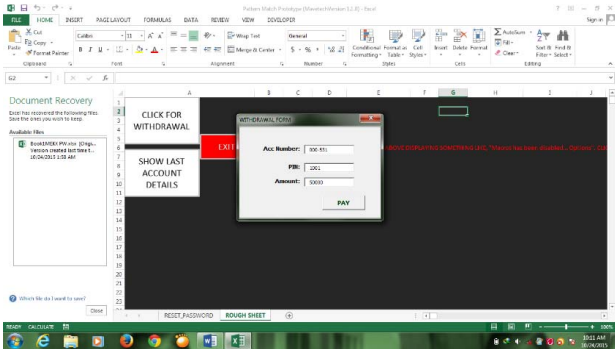


Figure 5 Display of debit above threshold limit

In a situation whereby the imposter tries to withdraw N50,000.00 he would be faced with another authentication question created by the IT ADMIN group and which was supplied by the customer. In the case of Mr. Anyanwu his withdrawal pattern as matched on the database does not exceed N5,000.00. Immediately the imposter enters an amount that doesn’t match the withdrawal pattern say ‘N50,000.00’, the system prompts for additional authentication, as shown in Figure 6, in this case the system requests for the name of his ELDEST NIECE.

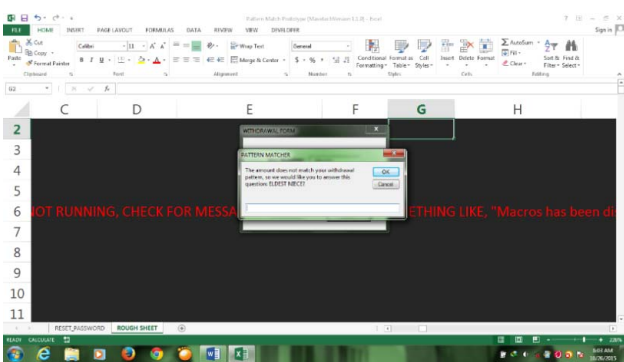


Figure 6 Display of pop up message as customer has exceeded threshold limit

On supplying the wrong name say ‘CHIOMA’, the system will match it to the already provided answer on the database.

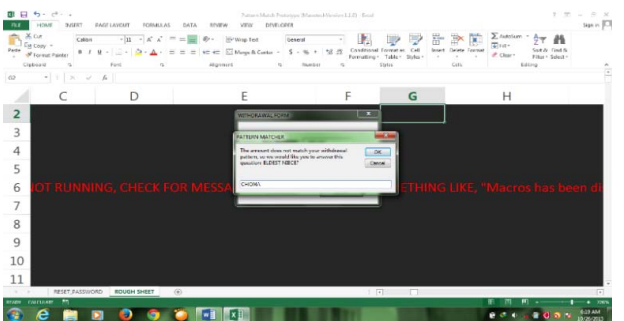


Figure 7 Password inputted as limit is exceeded.

On returning false, an error message is displayed which says ‘WRONG TRY AGAIN’.

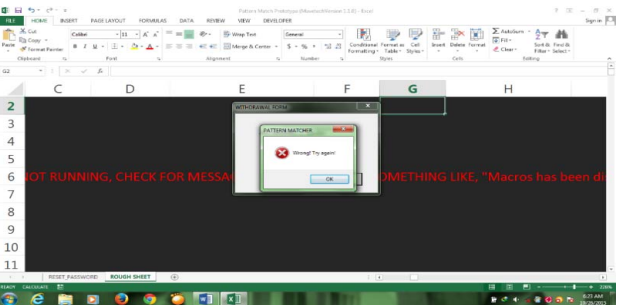


Figure 8 Display of error message as a wrong password was inputted.

Consistent trial of wrong answers will definitely put the paying teller on enquiry and definitely the security team would be invited as shown in Figure 7, Figure 8 and Figure 9. But if it happens on the ATM machine, the machine has no alternative than to trap the ATM card being used by the fraudster. Another important fact to note is that the answer to be given is also case sensitive, so it doesn’t end in knowing the name of Anyanwu’s eldest niece, the case pattern is also very necessary. For stronger authentication it can be a mixture of lower and upper case, or special character provided that the account owner can remember the sequence.

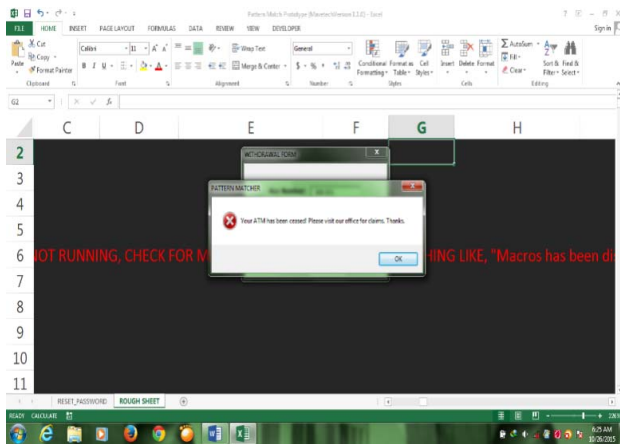


Figure 9 ATM trapped as penalty for inputting wrong password twice

In other words, the database is updated on a regular basis as transactions are consummated.

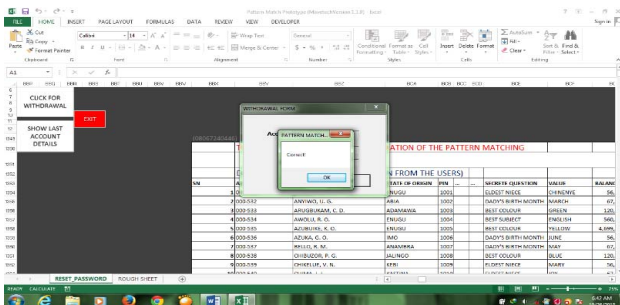


Figure 10 Correct password inputted

On inputting the right answer to the question, the system immediately acknowledges that it matches the answer in the database, debits the account and displays CORRECT, Figure 10. The system is therefore configured with these same parameters to be observed for every other customer of the bank. The design memorizes details of customer's transactions in the bank, the withdrawal branch and points for ATM and the keeps track of their account balances. Upon exceeding the accustomed range of amount or withdrawing from a new branch/point a further authentication is required. By so doing, transaction cannot be consummated until authentication is passed.

A. Risk Assessment/Loan Disbursement

Another area of application of data mining techniques to reduce fraud or detect it before it happens is the area of risk assessment/loan disbursement. This happens to be one of the core duties of banks; giving loans to help encourage and fund SME's. But how do they go about ascertaining who to disburse such marked out funds to and how can they ascertain the beneficiary that may default or not. In risk assessment, classification and association rule where applied to enable us predict the set/kind of people that can default in paying up loans.

Table 1 shown data of loan beneficiary from a financial institution, of which some paid up while others defaulted. It is to serve as a guide in disbursing loan to customers in the nearest future

The result of the analysis is as shown below:

SINGLE: It is observed that out of 228 'single's that took loan from the bank that period, only 152[66.67%] had a GOOD record while 76[33.33%] had a BAD record. The support [i.e the frequency or number of times that it occurred] that Singles have Good credit standing is 35.9% while the confidence [i.e. the number of times that the statement is true] that a Single will always have a good credit standing is 66.67%.

DIVORCED: It is observed that out of 138 divorced people that were in the survey, 20 of them [14.5%] recorded good credit standing while 118[85.5%] had a record of bad credit standing. The support [i.e. the frequency or number of times that it occurred] that Divorcees have Good credit standing is 4.7% while the confidence [i.e. the number of times that the statement is true] that a Divorces will always have a good credit standing is 14.5%.

MARRIED: It is observed that out of 58 Married people that were in the survey, 38 of them [65.5%] recorded good credit standing while 20[34.5%] had a record of bad credit standing. The support [ie the frequency or number of times that it occurred] that Married people have Good credit standing is 8.96% while the confidence [ie the number of times that the statement is true] that a Married person will always have a good credit standing is 65.52%.

From the above analysis we can predict that it is only safer and more advisable giving loans to individuals that are SINGLE.

V. RESEARCH RESULT

The design in section 4.1 memorizes details of customer's transactions in the bank, the withdrawal branch for cash transactions and points/locations for ATM transactions and then keeps track of their account balances. Upon exceeding the accustomed range of amount or withdrawing from a new branch/point a further authentication is required. By so doing, transactions cannot be consummated until authentication is passed.

The design in section 4.2 memorizes details of customer's loan transactions in banks (nationwide) with their status and their CREDIT STATUS/CONFIDENCE. Following the outcome of the analysis, the confidence of the institution disbursing the facility is determined to a large extent.

Fraud detection in real time is a highly demanding task that requires all the five [8] senses with the 6th sense inclusive. It requires all the parties involved to pay full and proper attention to details regarding every transaction, and also knowing their customers very well.

Implementation of this work can go a great mile is creating the awareness that people [fraudster] can go a great extent to take what does not belong to them and so therefore in order not to fall victim, we are hereby warned and advised to keep secrete all that pertains to our financial activities from bank account signature to ATM card details, hours of transactions, favorite branches, favorite ATM machine, etc.

TABLE 1 SHOWN DATA OF LOAN BENEFICIARY FROM A FINANCIAL INSTITUTION

checking acct	credit history	purpose	employ	Gender	Marital St	Housing	Job	Age	Credit Stai
OBalance	Current	Small Appliance	Short	M	Single	Own	Unskilled	23	Good
OBalance	Current	Furniture	Unemploy	M	Divorced	Own	Skilled	32	Bad
No Acct	Bank Paid	Car New	Long	M	Single	Own	Managem	38	Bad
Low	Current	Furniture	Short	M	Single	Own	Unskilled	36	Bad
Low	Delay	Education	Medium	M	Single	Rent	Skilled	31	Good
No Acct	Critical	Furniture	Short	M	Married	Own	Skilled	25	Good
OBalance	Current	Car New	Short	M	Married	Own	Unskilled	26	Good
High	Critical	Business	VeryShort	M	Single	Own	Unskilled	27	Good
No Acct	Current	Small Appliance	Short	M	Single	Own	Skilled	25	Bad
No Acct	Current	Small Appliance	VeryShort	F	Divorced	Own	Skilled	43	Bad
No Acct	Current	Business	Unemploy	M	Single	Rent	Managem	32	Bad
Low	Current	Car New	Short	M	Single	Rent	Unskilled	34	Good
OBalance	Current	Business	Short	M	Married	Own	Skilled	26	Good
Low	Current	Car New	Long	M	Single	Own	Skilled	44	Bad
Low	Critical	Car New	Medium	M	Single	Own	Unskilled	46	Good
No Acct	Current	Car Used	Short	M	Divorced	Own	Managem	39	Good
Low	Current	Furniture	VeryShort	F	Divorced	Own	Skilled	25	Bad
Low	Current	Car New	Medium	M	Single	Own	Skilled	31	Good
No Acct	Current	Repairs	Long	M	Single	Own	Skilled	47	Good
Low	Bank Paid	Education	VeryShort	F	Divorced	Rent	Skilled	23	Bad
OBalance	Critical	Furniture	Short	F	Divorced	Own	Skilled	22	Bad
Low	Current	Furniture	VeryShort	F	Divorced	Rent	Skilled	26	Bad
OBalance	Current	Furniture	Short	M	Married	Own	Skilled	19	Bad
No Acct	Current	Furniture	Short	F	Divorced	Own	Managem	27	Bad
No Acct	Current	Car New	Short	M	Single	Rent	Unskilled	39	Good
Low	Critical	Business	Medium	M	Single	Own	Skilled	26	Good
OBalance	Current	Car Used	Long	M	Single	Other	Skilled	50	Bad
No Acct	Critical	Car Used	Medium	M	Single	Other	Skilled	34	Good
Low	Current	Small Appliance	Medium	M	Single	Rent	Skilled	23	Good

TABLE 2 THE ANALYSIS OF LOAN MANAGEMENT

Rule	Count	Support	Confidence
Single	228		
Single => good	152	0.358491	0.666667
Single => bad	76	0.179245	0.333333
Divorced	138		
Divorced => good	20	0.04717	0.144928
Divorced => bad	118	0.278302	0.855072
Married	58		
Married => good	38	0.089623	0.655172
Married => bad	20	0.04717	0.344828

VI. RECOMMENDATION

- The real time fraud detecting proposal will save the banks from huge lose.
- It will save the customers from financial loss as well.
- It will build up people's confidence on keeping their funds in banks.
- It will make the society a better place with reduction in financial losses.
- It will discourage fraudsters from continuing in the act.

From the result obtained in the research work, the perfect means of authentication remains with biometrics.

In the area of risk assessment, it is not really a clear cut rule that two customers who happen to possess same qualities or attributes will definitely behave same, but in most cases (more than 65%) the behaviors can be the same.

VII. REFERENCES

- [1]. NIBSS 2015 www.nibss-plc.com.ng
- [2]. kim, Rachel and Monathan, Mary Javelin Strategy and Research. 2008 Identity Fraud Survey Report. February 2008.
- [3]. The Nilson Report. July 2007
- [4]. Data Mining Techniques and Their Implementation in Blood Bank Sector(Ankit Bhardwaj, Arvind Ssharma, V. K. Shrivastava/International Journal of Engineering Research and Applications (IJERA) vol2, Issue4 July-August 2012, pp 1304
- [5]. <http://www.alaric.com/solution/fraud-detection>
- [6]. <http://www.fico.com/en/blogs/analytics-optimization>.
- [7]. Hillol Kargupta, Anupam Joshi, Krishnamoorthy Siva Kumar, Yelena Yesha,"Data Mining:Next Generation Challenges and Future Directions",2005.
- [8]. Litan, Avivah. Gartner Research. Bank Spending on Fraud and Authentication Rises, but Not Due to Red Flag Regulations. May 2008.