Real-Time Detection of Fraudulent Transactions in Retail Banking using Data Mining Techniques

M. Mubasher Hassan

Department of Information Technology and Engineering (ITE)
Baba Ghulam Shah Badshah Badshah University (BGSBU), Rajouri (J&K), India
mubasher2003@gmail.com

Tabasum M.

Department of Computer Science, School Education
Government of Jammu and Kashmir, J&K, India
tabasum.mirza@gmail.com

***Abstract*— This paper presents one of the approaches of data mining for detection of fraudulent transactions in retail banking in real time scenario. The fraudulent transactions should be detected at real time out of enormous number of retail transactions being carried out simultaneously by the bank and blocked by the system to prevent the losses before they occur, sent for verification at a higher level where its authenticity should be proved or otherwise the transaction will be blocked from execution and further action will be carried out like reporting of transaction etc.**

Keywords - real-time fraud detection; Bayes classification; decision trees; data mining.

I. INTRODUCTION

All standard Banking transactions have crossed the physical boundaries of the bank where verifying authenticity and authorization of every transaction by bank is not possible. Retail banking transactions are growing enormously in volume and electronic remittances and transfers have changed the perspective of banking. Providing ease of access this anywhere anytime banking poses security risks also. Electronic payment systems are being extensively exploited for purpose of money laundering. Financial organizations around the world are losing approx. 5% of the annual revenue to frauds. Detection of frauds before they occur remain as main concern for 78% retail banks worldwide [1]. These frauds make banks liable to reputational and financial losses and may result in losing customers because of trust deficit created between the customer and bank.

Fraudulent transactions contribute to small percentage of total transactions but detecting that small percentage out of bulk of routine transactions is a difficult task without affecting the accessibility of global anywhere banking. Fraud detection should be done in real time to prevent losses by blocking attempted fraudulent transactions before they are executed. Consideration leads to demand of environment in which speed and accuracy in decision making is of utmost importance which can be done by analysis and comparison of behavioral patterns of customers in terms of account transactions.

Enforcing constraints on number of transactions or transaction amount in accounts will not solve the problem rather create problems for customers in terms of ease of access as retail transactions in accounts are unpredictable and unbalanced. Threshold amount associated with customer profiles can be good indicators but cannot detect frauds occurring within the threshold limits. Categorization of customer accounts into variants cannot resolve the issue as well as generalizing c needs of individual customers belonging to a particular variant group like savings account, pension account, current account, student account is not feasible and will not be same for different customers. Their transaction patterns shows remarkable differences and also vary with different socio-economic factors like variable sources of income, expenditure etc. Transactions in accounts are also time variant like for salaried class more transactions occur in the beginning of the month and shows decline in later days of month. Transactions in customer accounts can increase during festive seasons or business customers are most likely to withdraw more money on some occasions and deposit more amount on other occasions depending upon flow of money in business etc. Other factors like age, marital status, and number of dependents, additional sources of income, expenditure, inflation rates, and geographic location of account holder also effect the transaction amounts as well as no. of transactions in accounts.

II. METHODOLOGY ADOPTED

Banking transactions generate tremendous amount of data every day in a continuous manner. This data can be using for analytic purpose to discover knowledge that is useful for varying purposes. Data mining techniques are already popular in various banking applications like loan default prediction, credit risk assessment, credit scoring, credit card fraud detection etc.[2]. For the purpose of detection of fraudulent transactions in retail banking we can use data mining techniques to analyze transaction pattern of each customer and detecting suspicious fraudulent transaction attempts[3].

Retail transactions like cash deposit, cash withdrawal and transfer history of the each individual customer generally follow a particular pattern with occasional deviations. In real time fraud detection the suspicious fraudulent transaction should be detected, blocked and then the sent for verification at a level higher than normal to authorize it as legitimate transaction. In this additional level of scrutiny, if the transaction is proved as secure transaction it will be carried out otherwise fraudulent transaction will be blocked and exception will be generated for further necessary action like reporting of transaction etc.

The methodology for fraudulent transaction detection works on the fact that fraudsters or money launderers carry out high value transactions i.e. transactions of high amount or large no. of small transactions in accounts. So, when no. of transactions outnumber the average no. of transactions in an account or when transaction amount exceeds the average transaction amount or both the transaction is considered as suspicious of being fraudulent transaction. These averages can be calculated by analyzing transaction history pattern of the customer and combined with other parameters for classification of transactions.

III. DATA MINING

Data mining, also known as knowledge discovery is the process of analyzing and exploring large sets of data in order to discover meaningful patterns and rules. This deduced knowledge is useful in different applications.

Data mining is used to extract knowledge from large volumes of data. Data mining algorithms use the large amounts of data to build and train the models to perform the data mining tasks.

A. Data mining tasks

- Classification
- Estimation
- Prediction
- Association
- Clustering
- Profiling [4]

Use Decision Trees is simple and intuitive supervised learning technique of data mining used for predictive modeling. This non parametric method is widely used for classification and handles mixed data types, requires less data cleaning in comparison to other algorithms. Decision trees use top down approach and partition data into subsets based on similar attributes used to create training model which can predict class of target variables by learning decision rules from training data sets[5].

B. Bayes Classification

Bayes Classification also known as probabilistic networks or belief networks is predictive classification algorithm based on Bayes theorem by Thomas Bayes. This supervised classification technique is used for predicting class of unknown data by performs probabilistic predicts i.e. predicts class membership probabilities.

Input training data set consists of records and each record has a class label .model is used to classify test data for which class label is unknown.[5] Bayesian networks are well suited for real time prediction. Bayesian classifiers exhibit high accuracy and speed when applied to large data and works well with multiclass predictions [6].

Bayesian probability of class:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Where $P(c|x)$ is posterior probability

$P(x|c)$ is likelihood

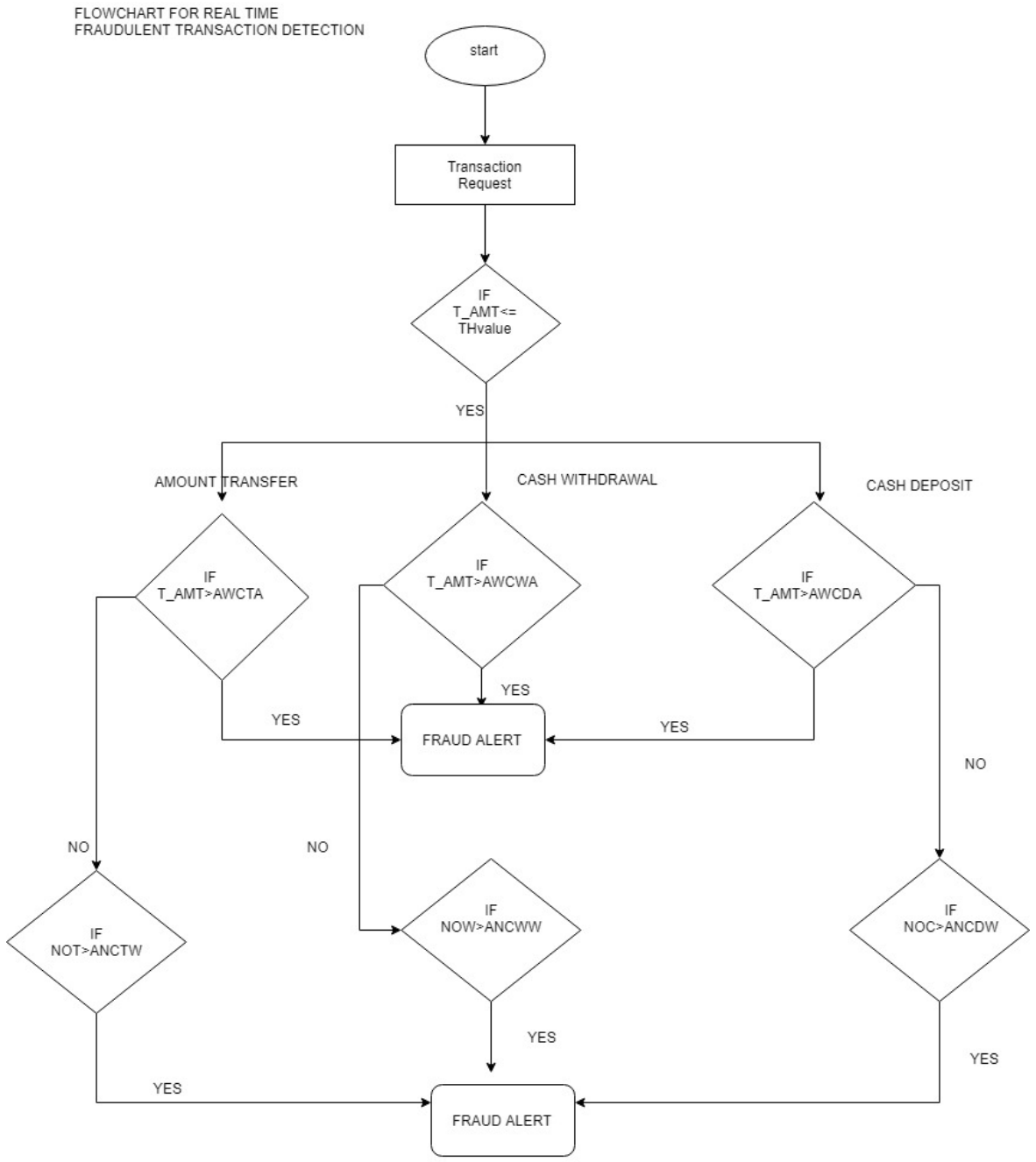
P(c) is class prior probability

P(x) is predictor prior probability

$P(c|x)$ = P(X1|c)x P(X2|c)x P(X3|c)X……x P(Xn|c)x P(c)

IV. CONCEPTUAL FRAMEWORK

A. Flowchart for real time fraudulent detection



B. Variables used

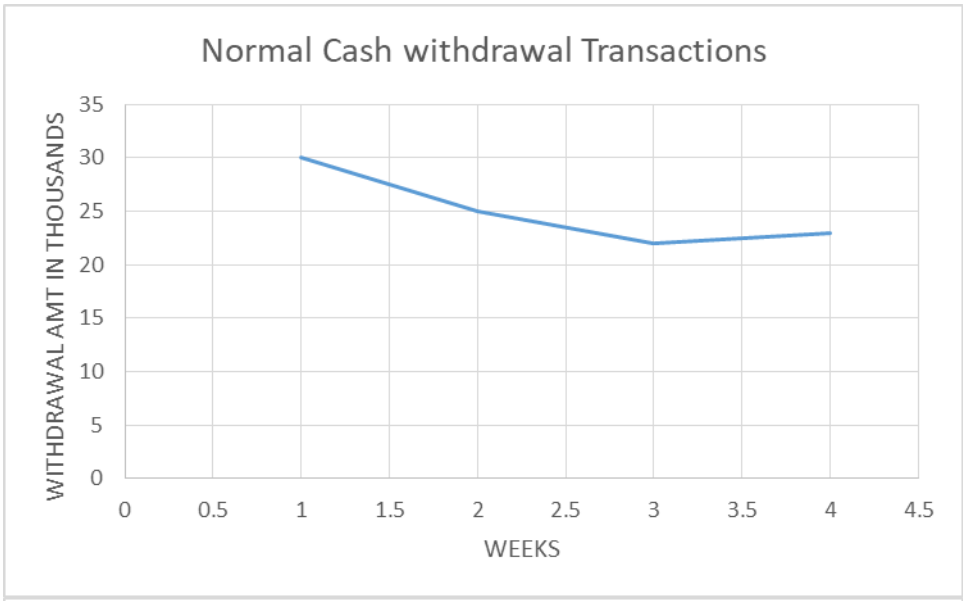
T_AMT=TRANSACTION AMOUNT
TH_VALUE=THRESHHOLD AMOUNT
NOT=NO. OF TRANSFERS
NOW=NO. OF WITHDRAWALS
NOD=NO. OF DEPOSITS
ANCTW=AVERAGE NO. OF CASH TRANSFERS PER WEEK
ANCDW= AVERAGE NO. OF CASH DEPOSITS PER WEEK
ANCWW= AVERAGE NO. OF CASH WITHDRAWALS PER WEEK
AWCTA=AVERAGE WEEKLY CASH TRANSFER AMOUNT
AWCWA= AVERAGE WEEKLY CASH WITHDRAWAL AMOUNT
AWCDA= AVERAGE WEEKLY CASH DEPOSIT AMOUNT

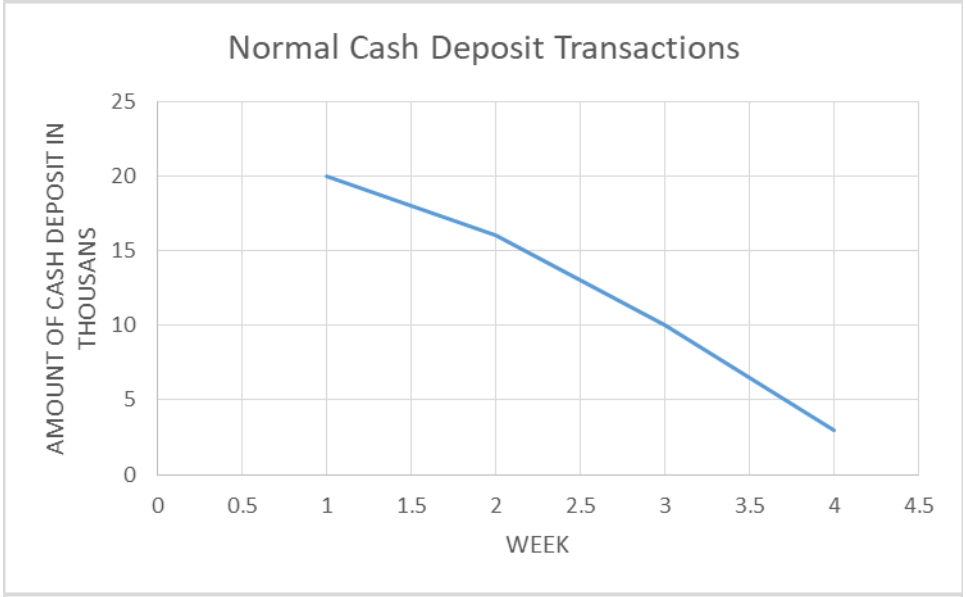
C. Algorithm for fraud detection

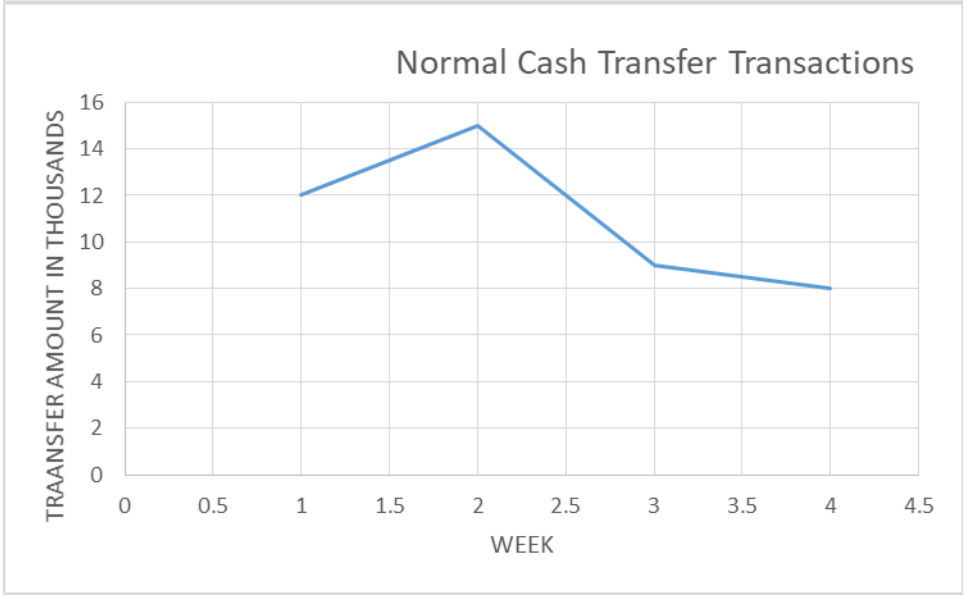
START
TRANSACTION ENTERS THE SYSTEM
IF T_AMT<=TH_VALUE
IF TRANSACTION_TYPE==CASH DEPOSIT GO TO DEPOSIT
ELSE IF TRANSACTION TYPE==CASH TRANSFER GO TO TRANSFER
ELSE IF TRANSACTION TYPE==CASH WITHDRAWAL
GO TO WITHDRAWAL
IF T_AMT>AWCDA
FRAUD ALERT
ELSE IF NOD>ANOCDW
FRAUD ALERT
IF T_AMT>AWCTA
FRAUD ALERT
ELSE IF NOT>ANOCTW
FRAUD ALERT
IF T_AMT>AWCWA
FRAUD ALERT
ELSE IF NOW>ANOCWW
FRAUD ALERT

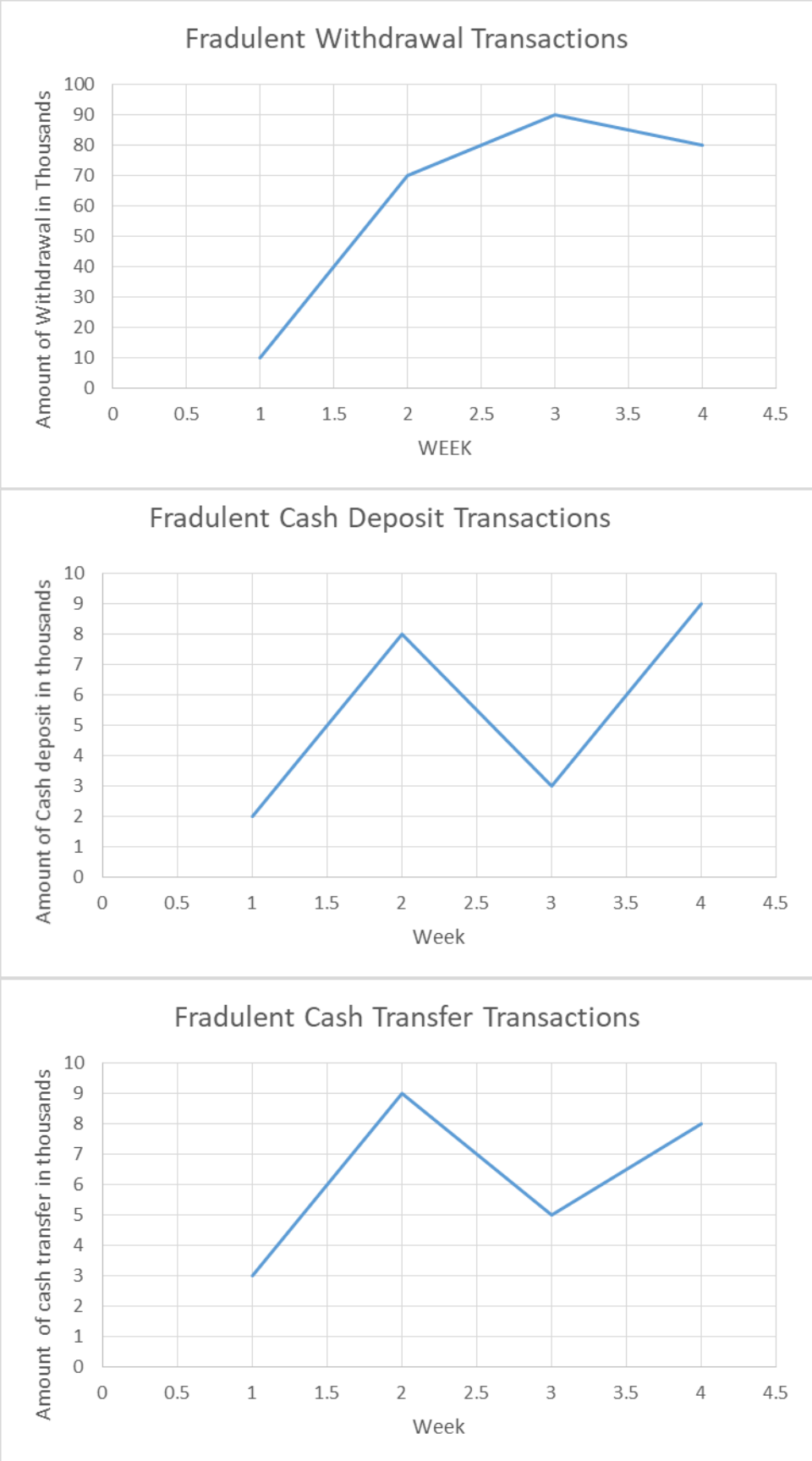
The fraud detection works by classifying the transaction into fraudulent transaction based on above mentioned parameters along with other parameters as shown in flowchart i.e. if no. of cash deposit transactions in an account exceeds its average no. of cash deposit transactions per week or if transaction amount exceeds its average cash deposit transaction amount the transaction is suspicious of fraudulent transaction.[7] In this case fraud alert will be generated and transaction needs another level of verification to be considered as legitimate transaction for execution otherwise the transaction will be blocked and reported. We are using decision trees and Bayes classification for this purpose[8].

D. Graphical representation of genuine and fraudulent transactions in a customer account.








Fradulent Withdrawal Transactions
Amount of Withdrawal in Thousands
100
90
80
70
60
50
40
30
20
10
0
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
WEEK
Fradulent Cash Deposit Transactions
Amount of Cash deposit in thousands
10
9
8
7
6
5
4
3
2
1
0
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
Week
Fradulent Cash Transfer Transactions
Amount of cash transfer in thousands
10
9
8
7
6
5
4
3
2
1
0
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
Week

V. CONCLUSION

Real time fraudulent transaction detection based on historic transaction pattern of individual customers can help in detecting frauds before they actually cause damage and stop such transactions from execution thereby reducing financial and reputational losses to banks. Challenge is to reduce false positives as the customer account transactions show timely variations and can be unbalanced depending upon various personal, sociological and economic factors. Detection models should continuously upgrade themselves, learning algorithms should be used and global events related to fraud detection should be considered for knowledge discovery and new emerging situations should be handled.

REFERENCES

[1] E. Group, "Fraud Analytics in Retail Banking," pp. 1–4, 2014.
[2] B. Rajdeepa and D. Nandhitha, "Fraud Detection in Banking Sector using Data mining," vol. 4, no. 7, pp. 2013–2016, 2015.
[3] B. Rajdeepa et al., Realtime Fraud Detection in the Banking Sector Using Data Mining Techniques/Algorithm, vol. 4, no. January. Elsevier B.V., 2016.
[4] M. J. A. Berry and G. S. Linoff, AM. .
[5] C. I. Mining, "Techniques in Data Mining : Decision Trees Classi cation and Techniques in Data Mining : Decision Trees Classi cation and Constraint-based Itemsets Mining," 2001.
[6] "Data Mining Algorithms for Classification," no. January, 2008.
[7] J. West, M. Bhattacharya, and R. Islam, "Intelligent Financial Fraud Detection Practices : An Investigation."
[8] C. Phua, V. Lee, K. Smith, and R. Gayler, "A Comprehensive Survey of Data Mining-based Fraud Detection Research."

AUTHORS PROFILE

Malik Mubasher Hassan: Received his B.Tech degree from University of Jammu followed by M.Tech degree from NIT Srinagar in 2007. He is presently working as faculty in the Department of Information Technology and Engineerong (ITE) at Baba Ghulam Shah Badshah University Rajouri (J&K), India-185234. His specialization is in wireless communication, optical wireless, computer Networks and Cloud Computing.

Tabasum Mirza: She has receceived her BCA and MCA from University of Kashmir in 2008. She is presently working as Lecturer in the Department of Computer Science School Education, Government of Jammu and Kashmir, India. She has a 6.5 years experience of working in JK Bank Pvt. Ltd. Her specialization are software Engineering, Java Prograamming and Data Mining.