International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Internet Banking Fraud Detection using HMM and BLAST-SSAHA Hybridization

Avanti H. Vaidya[1] , S. W. Mohod[2]

[1, 2]Computer Science and Engineering Department, R.T.M.N.U University, B.D. College of Engineering Wardha, India

Abstract: ***With the rise and swift growth of E-Commerce, credit card uses for online purchases has increased dramatically and it caused sudden outbreak in the credit card fraud. Fraud is one of the major ethical issues in the credit card industry. With both online as well as regular purchase, credit card becomes the most popular mode of payment with cases of fraud associated with it are also increasing. For this purpose an efficient fraud detection system is necessary. This paper presents the detection of fraud transaction using BLAST-SSAHA Hybridization which is used for the optimization of dataset and Hidden Markov Model. At the same time we have tried to ensure that genuine transaction are not rejected by making use of one time password that was generated by bank server and sent to genuine cardholder.***

Keywords: Internet Banking, Hidden Markov Model, BLAST-SSAHA Hybridization

1. Introduction

Data mining is a powerful tool to help organizations to extract hidden predictive information from large databases. The various data mining tools allowing business to make proactive, knowledge driven decisions. Online banking service is the most popular and provides a fast and easy way to make transaction. Internet banking has their separate account for users. It is managed by banks or retail store. Net banking is a process over the internet to make the banking process effectively. The bank has automatically updates the customer accounts and records. Transactional and non transactional features provided by online banking are as follows.

- Transactional
  1. Funds transfer between two customers
  2. Paying third parties
  3. Investment purchase or sale
  4. Loan applications
- Non transactional
  1. Viewing account balance
  2. Viewing recent transaction
  3. Ordering cheque book
- Supports transaction approval process.
- Wire transfer.
- Financial institution administration.
- Support of multiple users with authority.

A. BLAST-SSAHA Hybridization

BLAST is used to determine similarity of incoming sequence of transactions with the genuine card holders while SSAHA is used to give good results of alignment of long sequences. BLAST consists of three steps. In first step, it compiles a list of high-scoring words from given query sequence. In second step each high scoring word is compared with the database sequencs and if it is identical to a word in a database, a hit is recorded. In third step every hit sequence is extended and the extension is stopped as soon as the similarity score becomes less than the threshould value. All the extended segment pairs whose scores are equal to or greater than the theshould are retained are called as High-scoring segment pairs (HSPs). SSAHA is a two stage algorithm. In the first stage, a hash table is constructed from sequences in the database and in the second stage query words are searched from hash table.

B. Hidden Markov Model

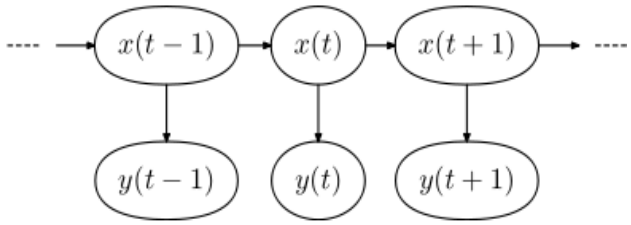

Figure 1: Architecture of Hidden Markov model

Fig 1 shows a general architecture of hidden markov model. Each oval shape represents a random variable that can adopt any number of values. The x(t) and y(t) represents the random variables. The x(t) is hidden state and y(t) is a observation at time 't'. The markov property states that the hidden variables x(t) at all times depends on the values of the hidden variables x(t-1). A hidden markov models are one of the most popular models in machine learning and provides statistics for modeling sequences. A probability distribution over sequences of observations is defined by using HMM. HMM maintains a log of several user transactions which provides a proof for the bank. HMM reduces substantial work of an employee since it maintains a log.

2. Literature Survey

Sushimito Ghosh and Douglas L. Reilly in (1994) have worked on neural network based fraud detection system. Using data from a credit card issuer, a neural network based fraud detection system was trained on a large sample of labeled credit card account transaction and tested on a holdout dataset that consisted of all account activity over a subsequent two months period of time. The network detected significantly more fraud accounts (an order of

Licensed Under Creative Commons Attribution CC BY

magnitude more) with significantly fever false positives (reduced by factor of 20) over rule based fraud detection procedures. [1].

Emin Aleskrov, Berned Freisleben and Bharat Rao in (1997) proposed a neural network based database mining system for credit card fraud detection which is identified as a CARDWATCH. The system is based on neural network learning module, provides an interface to a variety of commercial databases and has a comfortable graphical user interface. In case of modern corporate databases, copying huge data sets from database is not tolerable. The possibility to directly access different databases types becomes a critical requirement for modern database mining system. Therefore, the CARDWATCH system which is a more sophisticated yet straight forward graphical user interface has created. [2].

Mubeena Syeda, Yan-Qing Zhang and Yi Pan in (2002) proposed a work on Parallel granular neural networks for fast credit card fraud detection. A parallel granular neural network (GNN) is developed to speed up data mining process and knowledge discovery process for credit card fraud detection. The entire system is parallelized on Silicon Graphics Origin 2000, which is shared memory multiprocessor system consisting of 24 CPU, 4G main memory and 200 GB hard drive. Around eight scenarios are employed for detecting purpose [3].

Xuan Dau Hoang, Jiankun Hu, Peter Bertok in (2003) presented a new method to process sequences of system calls in order to detect anomaly intrusion. They proposed a Multi-Layer model for anomaly intrusion detection using program sequences of system calls using Hidden Markov Model. They performed their experiments on Unix Sendmail program have shown that the model is better in detecting anomalous behavior of program in terms of accuracy and response time. [4].

Vasilis Aggelis in (2006) proposed a work on offline internet banking fraud detection. They have demonstrated one successful fraud detection model. The main scope is to present its contribution in fast and reliable detection of strange transaction including fraudulent ones. The Offline internet banking fraud detection system offers many benefits to bank and to customers as well. New data is imported into the database in constant time frames, not in real time [5].

Osama Dandash, Phu Dung Le and Bala shrinivasan in (2007) has worked on internet banking models using a security analysis. They stated that internet banking fraud can be performed internally by genuine staff or externally by customers or suppliers. A security analysis of the proposed internet banking model compared with that of the current existing models used in fraudulent internet payment detection and prevention [6].

Abhinav Shrivastava, Amlan Kundu, Shamik Sural and Arun Majumdar in (2008) proposed an application of Hidden Markov Model (HMM) in credit card transaction processing. They have used the ranges of transaction amount as the observation symbols, whereas the types of item have been considered to be states of the HMM. They have suggested a method for finding the spending profile of cardholders [7].

Osama Dandash, Yuiling Wang , Phu Dung Le and Bala shrinivasan in (2008) have proposed an efficient new scheme used to prevent fraud by applying different security algorithms where hacking one secret will not compromise the whole system's security because the system does not rely on fixed values[8].

Qinghua Zhang in (2009) provided a survey on fraud risk prevention of online banks. Their aimed, in the first hand, at giving a discussion on the fraud risks of online banking, introducing the current application situation of information sharing mechanism in respect of internet fraud outside China as well as the development of such concept in China [9].

Amlan Kundu, Suvasini Panigrahi, Shamik Sural and Arum K. Majumdar (2009) have proposed to use two-stage sequence alignment in which a profile analyzer (PA) first determines the similarity of an incoming sequence of transactions on a given credit card with the genuine cardholder's past spending sequences. The unusual transactions traced by the profile analyzer are next passed on to a deviation analyzer (DA) for possible alignment with past fraudulent behavior [10].

Khyati Chaudhary and Bhawna Mallick in (2011) they investigated the factors and various techniques involved in credit card fraud detection during/after transaction as well. Efficient and well-organized credit card fraud detection system is a greatest requirement for any card issuing bank. Credit card fraud detection has drawn quite a lot of interest from the research community and a number of techniques have been proposed to counter/identify credit card fraud [11].

Sunil S. Mhamane and L.M.R.J Lobo in (2012) explained about how Fraud is detected and prevented using Hidden Markov Model. At the same time, they have tried to ensure that genuine transactions are not rejected by making use of one time password that was generated by the Bank server and sent to the particular customers through SMS to their mobile number which is registered in the system [12].

S. Esakkiraj, S. Chidambaram in (2013) designed a model with sequence of operations in online transactions by using Hidden Markov Model (HMM) and decides whether the user act as a normal user or fraud user. The HMM is initially trained with customer's last few transactions. In the trained system, the new transaction is evaluated with transition and observation probability, system finds the acceptance probability and decides the transaction will be declined or not. [13].

Licensed Under Creative Commons Attribution CC BY

3. Proposed System Process Flow

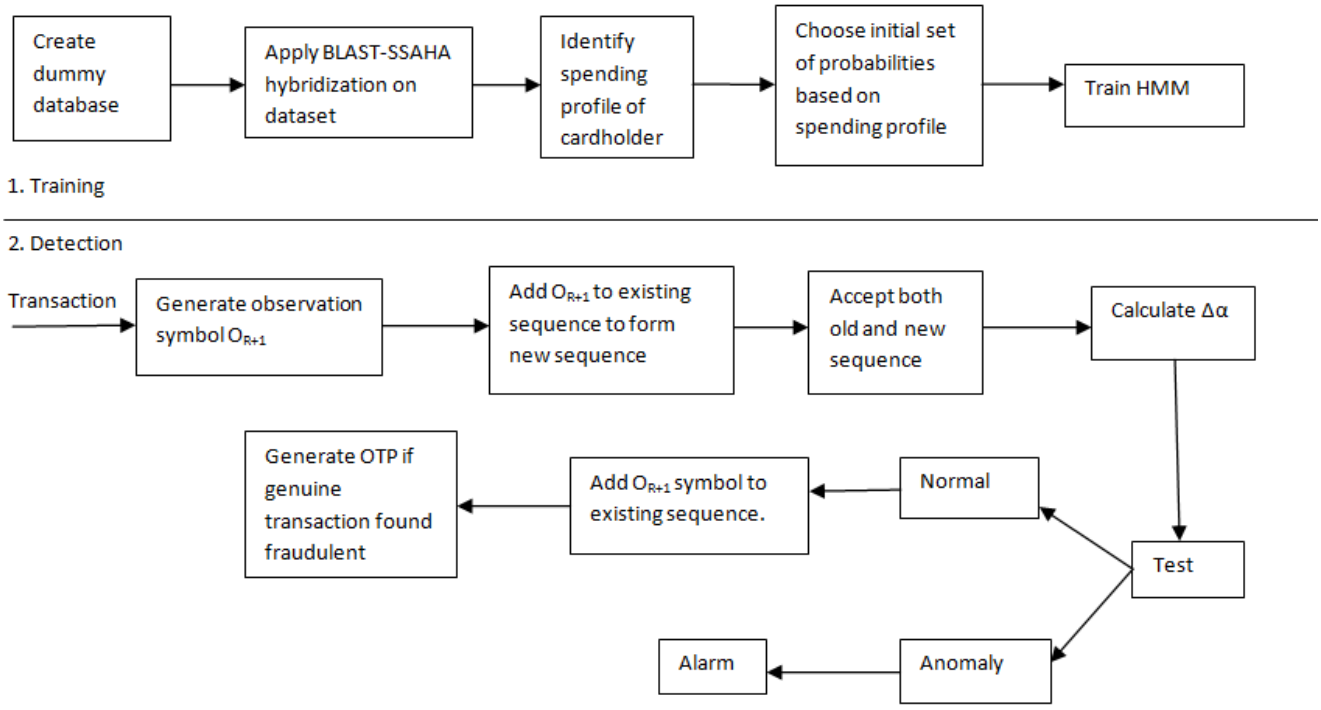

Figure 2: Fraud detection system process flow

The Fig2 describes the fraud detection system flow. The fraud detection system contains two phase named training phase and detection phase.

C. Training Phase

Dummy database is created because of privacy of banks. Dummy database consists of valid database and fraudulent database. The BLAST-SSAHA Hybridization is applied on a both the valid and fraudulent database for the optimization. The HSPs, concept of BLAST algorithm are generated as two pairs. The first pair contains credit card number, income and account balance while second pair contains credit card number, income and transaction. k-tuple table , concept of SSAHA is created to detect whether the generated HSPs is repeated or not in the database. The k-tuple table contains tuple weight and sequence index. Two k-tuples are called overlapping if they share same amount of information between them. Every distinct k-tuple is assigned an integer value which is called "tuple weight". Sequence index is used to represent i_{th} sequence of the data in the database. If the generated HSPs are repeated in the database then sequence index will hold the value more than one and if it is not repeated then sequence index represent value equals to 1. Then identify the spending profile of cardholder which contain transaction amount information and choose initial set of probabilities on spending profile of cardholder and trained HMM on the set of sequence.

D. Detection Phase

After the HMM parameters are learned, we have taken the symbols from a cardholder's training data and form an initial sequence of symbols. Let O_1 ,O_2 ,….O_R be one such sequence of length R. This recorded sequence is formed from the cardholder's transactions up to time t. We input this sequence to the HMM and compute the probability of acceptance by the HMM. Let the probability be α1, which can be written as:

$$\alpha 1 = P(O_1, O_2, \ldots O_R \mid \lambda)$$

Let O_{R+1} be the symbol generated by a new transaction at time t+1. To form another sequence of length R, drop O_1 and append O_{R+1} in that sequence , generating $O_2,O_3,\ldots O_R,O_{R+1}$ as the new sequence. Input this new sequence to HMM and calculate the probability of acceptance. Let the new probability be α2.

$$\alpha 2 = P(O_2, O_3, \ldots O_R, O_{R+1} \mid \lambda)$$

$$\Delta\alpha = \alpha 1 - \alpha 2$$

If $\Delta\alpha > 0$ it means that the new sequence is accepted by the HMM with low probability, and it could be a fraud.

If the system detects genuine transaction as a fraud then one time password will be sent to the genuine cardholder in order to process genuine transaction. If the fraudsters stole the credit card number and process the transaction by entering transaction amount at the same time the one time password will be sent to the genuine cardholder and fraudster will not be able to perform transaction, system will generate the alarm and transaction process will be declined. So, the genuine transaction will not be rejected during its transaction time.

4. Experimental Results and Analysis

In this section we show the performance of system, experimental results and analysis on various transaction amounts. Table 1 below shows the list of transaction with

Licensed Under Creative Commons Attribution CC BY

different transaction amount and different categories of purchasing type.

Table 1: List of all Transaction

Transaction No.	Category	Amount in Rs
1st	1	25
2nd	3	560
3rd	2	50
4th	4	600
5th	2	450
6th	2	300
7th	1	700
8th	3	70
9th	3	400
10th	1	335
11th	4	426
12th	2	220
13th	1	315
14th	3	360
15th	4	100

According to this table we propose:

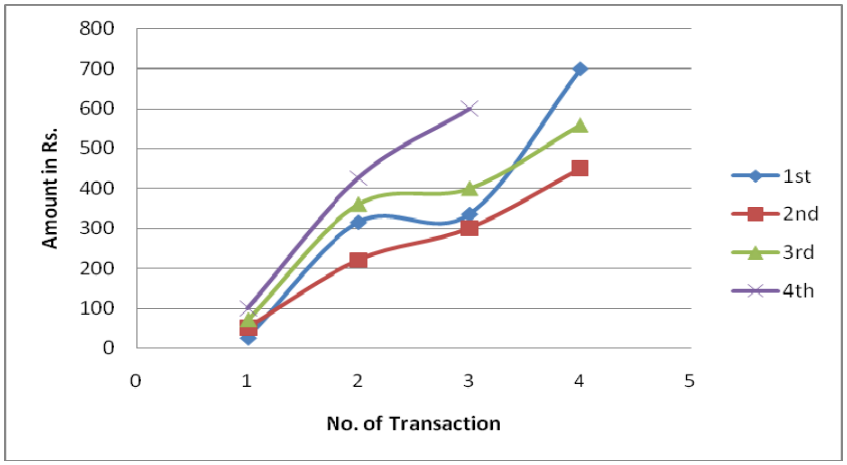

Figure 3: Transaction amount of each category.

Figure 4 shows the mean distribution of fraud transaction where the probability of fraud transaction is compared with the genuine transaction.

Licensed Under Creative Commons Attribution CC BY

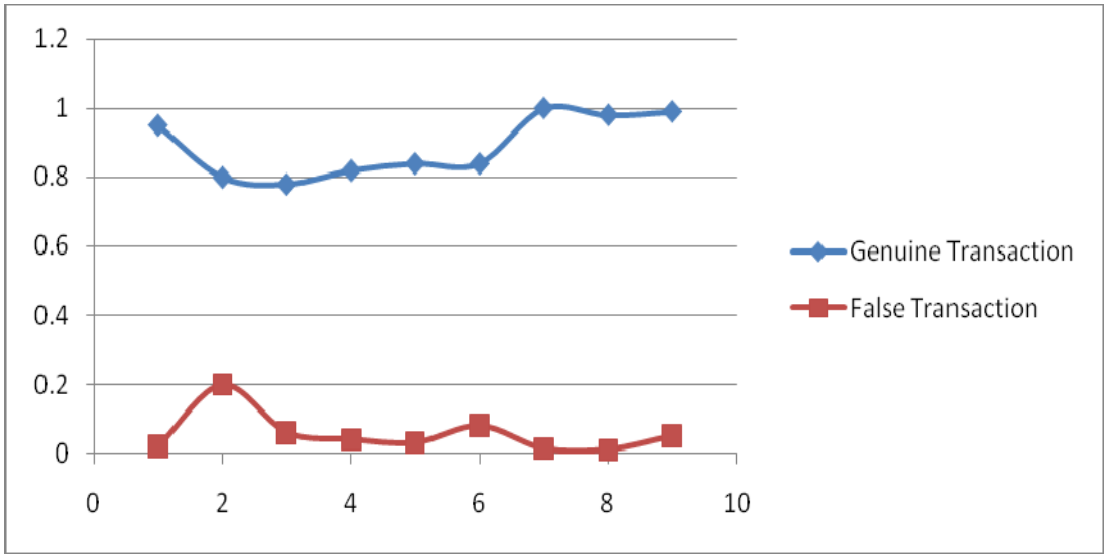

Figure 4: Mean distribution of Fraud Transaction

From the above figure it is noted that the when the probability of genuine transaction is going down correspondingly the probability of fraudulent transaction is going to increase and vice-versa. It helps to find out false alarm for the detection of fraud transaction.

Table 2: Experimental results of different range of transaction

Transaction No.	Transaction Amount(Rs.)	Transaction Found Genuine(G)/Fraudulent (F)	OTP Entered	Transaction Processed
1	250	G	-	YES
2	300	G	-	YES
3	450	G	-	YES
4	600	G	-	YES
5	700	G	-	YES
6	4000	F	YES	YES
7	5530	F	NO	NO
8	4300	F	YES	YES
9	345	G	-	YES
10	175	G	-	YES
11	2220	F	NO	NO
12	3350	F	YES	YES

Table 2 shows the performance of the system. The transaction amount called as observation symbol of HMM is taken into consideration. If the transaction amount is within the range of genuine cardholder then system detects the transaction as genuine transaction and allows processing the transaction of genuine cardholder. But if the genuine cardholder process transaction beyond the transaction amount range then system detects genuine transaction as a fraudulent transaction and at the same time system sends one time password (OTP) to genuine cardholder. If the cardholder enters the OTP which is generated by bank server correctly, the system continues his transaction as shown in Table 2 otherwise declined the transaction. On the other hand if there is a fraudster processing transaction of genuine cardholder beyond the range, again system sends OTP to genuine cardholder. Fraudster will not be able to see that OTP. If he tries to enter any random OTP then system produce message of declined transaction.

According to this we propose:

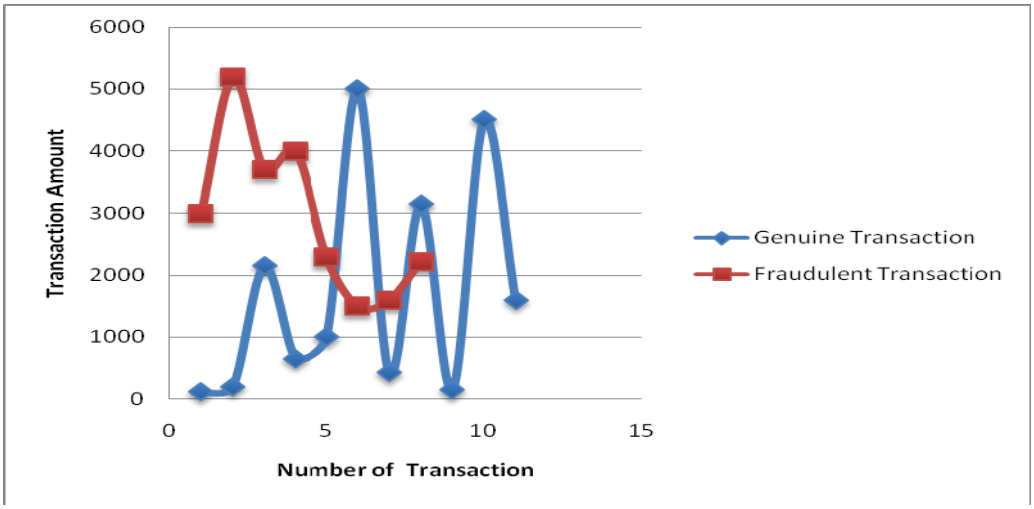

Figure 5: Comparison of genuine and fraudulent transaction with different ranges of transaction amount

Licensed Under Creative Commons Attribution CC BY

5. Future Scope

In my research transaction amount was considered as an observation symbol. In future we could take other parameters to make system strong and reliable. A different algorithm for checking fraud detection making system more and more accurate and reliable could be designed and implemented. Instead of HMM and BLAST-SSAHA Hybridization algorithm we could use another algorithm.

6. Conclusion

The efficient internet banking fraud detection system is an utmost requirement for any card issuing bank. We have proposed algorithm named BLAST-SSAHA Hybridization for optimization of dataset and HMM model for internet banking fraud detection. We have used transaction amount as the observation symbol. We have suggested a method for finding spending profile of cardholders as well as application of this knowledge in deciding the values of observation symbol and initial estimate of model parameters. We have also been explained how the HMM can detect whether the incoming transaction is a case of fraud or not. It also removes high false alarm rate generated by the system. We have generated One Time Password so that genuine transaction should not be rejected during transaction process.

References

[1] Ghosh and Reilly "credit card fraud detection with a neural network", IEEE Proceedings of the Twenty-Seventh Annual Hawaii International Conference on System Sciences, 1994.pp 621-630.
[2] Aleskerov, Emin, Bernd Freisleben, and Bharat Rao "CARDWATCH: A neural network based database mining system for credit card fraud detection.", Computational Intelligence for Financial Engineering. Piscataway, NJ: IEEE, 1997, pp.220-226.
[3] Mubeena Syeda, Yan-Qing Zbang and Yi Pan" Parallel Granular Neural Networks for Fast Credit Card Fraud Detection", IEEE Transaction .2002, pp.572-577
[4] X.D. Hoang, J. Hu, and P. Bertok, "A Multi-Layer Model for Anomaly Intrusion Detection Using Program Sequences of System Calls," Proc. 11th IEEE International Conf. Networks, 2003, Pp.531-536.
[5] Vasili s Aggelis "Offiine Internet Banking Fraud Detection". 0-7695-2567-9/06, 2006, IEEE Proceedings of the /first International Conference on Availability, Reliability and Security.
[6] Osama Dandash,Phu Dung Le and Bala Srinivasan " Security Analysis for Internet Banking Models". Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, IEEE transaction, 2007, pp. 1141-1146.
[7] Abhinav Srivastava, Amlan Kundu, Shamik Sural. "Credit Card Fraud Detection Using Hidden Markov Model" IEEE Transaction, January-March 2008. Pp. 37-47.
[8] Osama Dandash Yiling Wang and Phu Dung Leand Bala Srinivasan "Fraudulent Internet Banking Payments Prevention using Dynamic Key". "Journal Of Networks, Vol. 3, No. 1, January 2008".
[9] Qinghua Zhang "Study on Fraud Risk Prevention of Online Banks". International Conference on Networks Security, Wireless Communications and Trusted Computing.978-0-7695-3610-1/09, 2009 IEEE, pp 181-184.
[10] Amlan Kundu, Suvasini Panigrahi, Shamik Sural and Arum K. Majumdar"BLAST-SSAHA hybridization for credit card fraud detection", IEEE Transactions On Dependable And Secure Computing, Vol. 6, No. 4, October-December 2009. Pp. 309-315.
[11] Khyati Chaudhary and Bhawna Mallick "Exploration of Data Mining Techniques in Fraud Detection System", International Journal of Electronics and Computer Science Engineering 1765, ISSN- 2277-1956.
[12] Sunil S Mhamane and L.M.RJ Lobo "Internet Banking Fraud Detection Using HMM" , ICCCNT'12 26th_28th July 2012, Coimbatore, India. IEEE-20180.
[13] S. Esakkiraj and S Chidambaram "A predictive approach for fraud detection using Hidden Markov Model", International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 1, January- 2013, ISSN: 2278-0181.

Author Profile

Ms. Avanti Vaidya is a P.G. Scholar in Computer Science and Engineering from B. D. College of Engineering, Sewagram, MH, and India. She has received her BE degree from B. D. College of Engineering, Nagpur University, India in 2011. She has published three papers in International journal and presented two papers in International conference and one paper in national conference. Her research interest is Data Mining and warehousing, data security.

Prof. S. W. Mohod is working as an Assistant Professor (Sr.Gr.) in P.G. Department of Computer Science and Engineering at B. D. College of Engineering, Sevagram , MH, India. He did his B.E in 1995 and ME in 2006 in Computer Science and Engineering from Amravati university, Amravati. He has published many research papers in international journal and conference and national journal and confereces. He is a life member of ISTE and IE .His area of interest is data mining and networking.#

Licensed Under Creative Commons Attribution CC BY