

Improved Sheep Flock Heredity Algorithm Based Prevention of Credit Card Fraud Detection for Online and Offline Transaction

V.Mareeswari^{1*}, G. Gunasekaran²

^{1*}*Research Scholar, Department of Computer Science and Engineering, St.Peter's University, Avadi, Chennai, India, email: mareesh82.prasanna@gmail.com,*

²*Principal, Meenakshi College of Engineering, Chennai, India, email: gunaguru@yahoo.com.*

Abstract

One of the Intrusion Detection System is credit card fraud detection in data mining. The existing approaches validate the fraud occurrence by computing a communal analysis suspicion score of the credit applications. The key challenge of this paper is, to improve the efficiency of the credit card fraud detection method by verifying and validating the optimized parameters such as single and multiple attributes. The attributes of every application [offline/online] are verified using a newly developed procedure is **MLMA-[Multi-Level-Multi-Agent]** and it is verified all the attribute values are best one or not. For optimizing the attributes the **ISFH-[Improved Sheep Flock Heredity]** algorithm is used and those attributes are validated according to the time and response with optimal value. The experimental results of the proposed approach are compared with the existing approach results to compute the performance evaluation where the proposed approach experiments in DOTNET framework 2012 software.

Keyword

Intrusion Detection System, Credit Card Fraud Detection, Anomaly Detection, Data Stream Mining, Security.

Nomenclature

Symbol	Description
CC	Credit Card
CCD	Credit Card Detection
CFD	Credit Card Fraud Detection
GA	Genetic Algorithm
MLMA	Multi Level Multi Agent
ISFHA	Improved Sheep Flock Heredity Algorithm

BACKGROUND STUDY

One of the dishonor criminal acts in online banking is credit card fraud. One of the expensive identity crime applications is the credit card application. From the ethical point of view, the action taken against credit card fraud is the banks and credit card companies. But, the software companies try to provide a solution on behalf of the banks and the customer. The early systems are having limitations in terms of score computation and applying rules.

To overcome these limitations, the spike detection and communal detection are the two main processes applied to the existing paper. The communal detection methodology finds out the social relationships among the data inputs. The Spike detection finds out the duplicates in the input data to find out the attacks. In this paper, the application refers the credit card fraud detection in online transactions. The identity verification in this application is synthetic and real identity verification. This CFD application needs to deploy in various generalized-distributed applications like insurance, telecommunication, online transactions and so on. In abroad countries, all these kinds of applications uses a registered secret number as identity. It is well known that 75% of the world bank's run their business on mainframes. So there are chances of hacking the systems on online. IBM introduced some of the following techniques [1], where CFD application can follow for improving the detection accuracy such as:

- Identity Vulnerabilities
- Transaction Detection
- Workloads are Evaluated
- Remediation conductance
- Process Appeals

Several approaches and techniques were applied in the earlier researches, due to improve the detection accuracy and to provide high security. Sam Karl [2], proposed Bayesian classification with neural network [9] based approach for CFD. These approaches use the learning models and find out the fraud transactions. Kim and Kim [3] proposed an analysis where it combines both fraud and legitimate transactions to improve the comparison detection. Few scholars in the early stage used clustering methods for grouping fraud and fraud-less data [4], grouping patterns which are visible and invisible [5]. In general the clustering approach clusters the parameters in regions.

Foster & Stine [6] proposed a model which predicts the personal information from the bank data, for the user one who are using credit card alone. It verifies the non-linearity, missing values, standard errors, time taken for transaction and more transaction based destination addresses in the database. It is well known that the credit card based transaction is growing in the internet sales and purchases. In this case, the fraudsters make use of manipulating the credit card data in charge-back method [7]. Shilesh et al.[8] Utilized the hidden Markov model for analyzing the hidden entries of the credit card transaction in online payments. Ethics of banking is strongly provided for fraud detection [10]. Various types of credit card fraud in financial industries with the appropriate remedies are discussed in [11]. According to the survey given in [12], the Euro monitor International [13] says that 120 million numbers of credit cards were used in transaction in Germany alone. According to the increased number of credit card usage the fraud transaction is also getting increased. To detect the credit card fraud, the research scholars are proposing various approaches. In this paper, an optimized attribute based application approval is introduced for credit card delivery to a customer. In [15], utilized Genetic programming method for finding the fraud transactions. The motto of the existing approach was comparing the test data with the training data, where the training data were optimized by the GA. It uses principles of genetics and neural selection for solving the complex problems. The existing approach examines the results to detect the fraud between the credit card companies and their clients. In the existing systems, the transaction data are compared simply with the database data, but it in this paper, the pattern of the testing and training data are verified. Also, the existing system concentrates on the data provided after and during a transaction in online, which can detect the transaction is a fraud transaction after successful fund transfer. To overcome the issues in the existing system, in this paper, a prevention method is provided for detecting a card requester is a genuine person or not and it avoids providing a credit card to a fraud person.

EXISTING APPROACH

One of the security service company Max Mind [16], calculating a risk Score can determine the fraudulent. It uses the statistical analysis on IP-address, Devices, email-address, Geo location verification, proxy detection. Bank ID, and compare with the min Fraud network. It verifies the likelihood ratio of the available data and the input data of the card request. The range of the risk score is from 0.01 to 100. For example, if the request order has 20.00 risk Score is being a fraudulent.

Clifton Phua et al.,[17], proposed a multi layered detection approach for detecting credit card application fraud. Various existing approaches are non-data mining approaches compares the business rules and scorecards with the known fraud limitations. But, Clifton Phua et al., utilized Communal Detection and Spike Detection based fraud detection. The Spike detection finds out the duplicate, fraud data in dynamic attributes and the Communal detection find out the duplicate, fraud data in static attributes of the credit card applications. The combined CD and SD detect various attacks by comparing the input data with the persisted probe data. Since, the dynamic data may change according to the bank rules and increasing e-business, the detection rate and comparison time is poor. To improve the detection accuracy with in a stipulated time a MLMA approach is proposed in this paper.

MATERIALS AND METHODS

A real time data set is chosen for experimenting our approach to improve the efficiency in terms of detecting most recent fraud people applied for a credit card. The data are taken from [14], is a synthetic data having 50,000 credit card application information. In this data set most of the social attributes are very similar and twisted. The interval between the applications are in milliseconds. 75% of the attributes are treated as string attributes and other attributes are numeric. Some of the attributes are encrypted for privacy purpose. Each attribute of every application is filtered and verified for detection frauds. The number of fields in each record is 30 and the size of the data is 140 bytes. In the overall data, 20% of the data are showing fraud entries. Our proposed approach can evaluate the entire data ad classify the score to predict the normal and fraud data. The following section describes about the optimization process.

PROPOSED APPROACH

In this paper, MLMA model is taken for analyzing the CFD methods. There are two ways to apply credit card such as online and offline. Both online and offline

application data are fed into a software which can directly convert the data into separate fields in a table. MLMA approach is proposed in this paper in order to reduce the time complexity, improve the attribute verification accuracy and fraud detection accuracy. There are three agents *IAgent*, *BAgent*, and *CCAgent* are integrated in our approach. All the input data [attributes from the credit card application form] is read and separated into three categories as Personal, social and Official using an online/offline software. *IAgent* verifies the Personal information, *BAgent* verifies the Social information and the *CCAgent* verifies the official information obtained from the application form. It is assumed that, the three agents can behave as a civil software by comparing the input data [attributes] with the available data such as training data and real time data. The agents are also having permission to verify the fetched data formats and the values. The application is also having more credentials to be filled to get approved credit card for delivery.

Most of the fraud detection is obtained mostly from *BAgent* and *CCAgent* not from *IAgent*. *IAgent* always reads the encrypted data like DOB, Sex, Address, ID, and etc. But the *BAgent* compares the social information with the available real time information and finds the matching score. Similarly, the *CCAgent* compares official information with the available real time information and finds the matching score. Less cases, *IAgent* also give a less number of matching score where the same person may have a credit card with other banks or from other credit card companies. The meta-information, history, IP address of the system from where the input comes, the interval between the applications, and amount of credit are mainly focused and verified from the DB and a score is assigned for the analysis. After verification and scores assigned by the agents, all the attributes and relevant data values are fed into ISFH algorithm, and finds the optimal score for credit card approval. Also, the MLMA model used in this paper is shown in Figure-1.

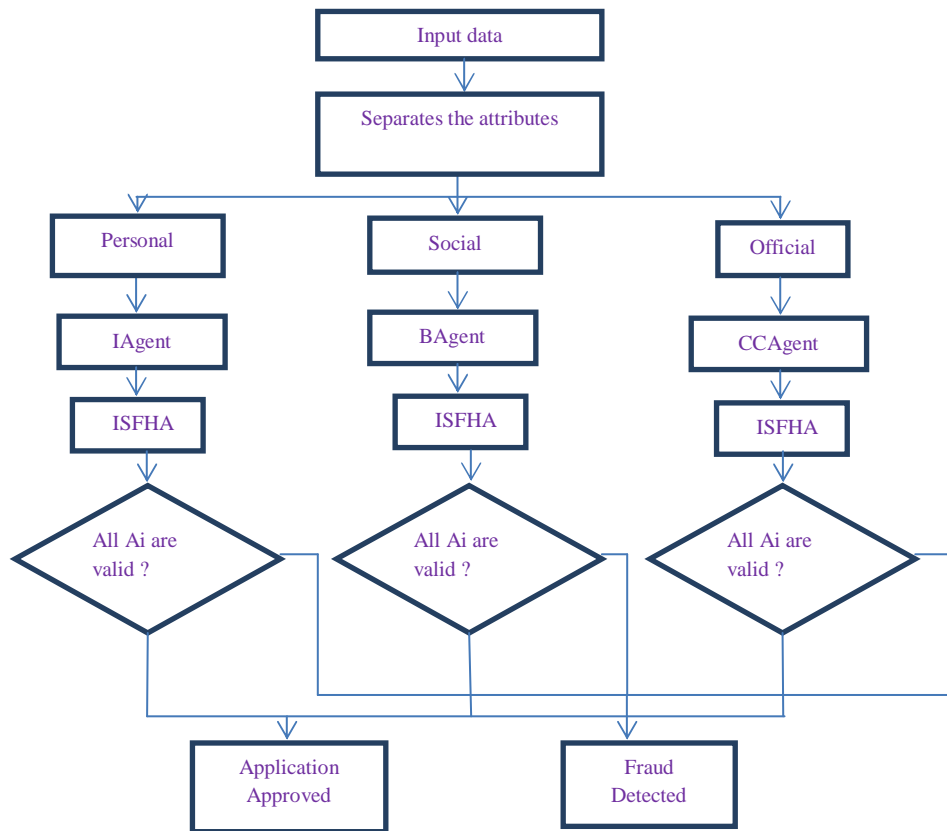


Figure-1: Multi Level Multi Agent Model for Analyzing the CFD analysis

IAgent fetches the personal data from the user application, **BAgent** fetches the bank information from the user as well as from the relevant bank, and the **CCAgent** fetches the credit card data from the user as well as from the credit card company for verification. All the fetched, verified data is optimized using ISFH algorithm and compare it with the training data. If the matched score is high, then the transaction is treated as genuine, else it is treated as fraud. ISFHA has a set of states where each state is connected with an optimized probability distribution. The set of probabilities is called as transition probabilities which are calculated on top of the transitions. Since, the number of state and observation are internal as well as external it should be optimized in each credit card request and which is compared to the recorded early transactions. Since the optimization technique is used, the detection rate of the false positive and false negative is reduced, because, the maliciousness is decided according to the FPR and FNR values, where the ISFH algorithm helps to predict the FPR and FNR.

In this paper, the CD can be obtained by comparing the present input values with the a correct value to be measured as genuine values.

$$\left(\begin{array}{l} \mathit{score} = 1 \\ \mathit{score} = 0 \end{array} \begin{array}{l} \sum_{i=1}^{i=k} \mathit{if} A_i == CV \\ \mathit{if} A_i \neq CV \end{array} \right)$$

Where, score=1 denotes that the value of the attributes [A_i], is matched with the correct value [CV] and score=0 denotes that the attributes are not meeting the perfectness. And the SD can be obtained by comparing the input values with the previous application values. The genuine values and the previous values are provided in the form of database to be compared to finding the matching score.

$$\left(\begin{array}{l} \mathit{score} = 1 \\ \mathit{score} = 0 \end{array} \begin{array}{l} \sum_{i=1}^{i=k} \mathit{if} A_i \geq PA_i \\ \mathit{if} A_i < PA_i \end{array} \right)$$

Where score=1 denotes it scored a value, if the attribute value [A_i], is greater than the previous attribute value [PA_i].

Developing a CFD application is very difficult due to some reasons, such as: 1) the companies do not share their databases, 2) the size of the database is growing day by day and 3) banking and companies are periodically changing the GUI on their applications. To address these issues, the models and the Metadata, and the software agents working intermediate are verified every time and grant permission only for the optimized entries. This paper proposes a CFD method with ISFH algorithm. ISFH algorithms are heuristic algorithms can provide better solutions within a stipulated time. At the time of credit card copied, theft, or captured by the fraud people it is usually applied up to its availability time is depleted. Therefore, even though several correctly classification methods are available, ISFH algorithm reduces the time, and improves the accuracy in finding the fraud by optimizing the parameters. The list of parameters used in this paper is given in the following Table-1.

Attributes	Description	Attributes
A1	Name	A18
A2	Middle Name	A19
A3	Last Name	A20
A4	Date of Birth	A21
A5	Gender	A22
A6	Qualification	A23
A7	PAN card Number	A24
A8	Email-ID	A25
A9	Mobile	A26
A10	Address Line-1	A27
A11	Address Line-2	A28
A12	City	A29
A13	Pin Code	A30
A14	Res. Number	A31
A15	Occupation Type	A32
A16	Bank of Credit Card	A33
A17	Last Transaction Date	A 34

Table-1: Attributes Taken in Sample Data

A set of sample personal, social and official attributes are given in the Table-1. The present values of the attributes are determined and a comparison between the dataset and the critical values based parameters is obtained for increasing the number of true alerts. To find out a better solution, ISFH algorithm is applied continuously to compute the critical values, frequency of the credit card, used location etc. shown in Table-1.

IMPROVED SHEEP FLOCK HEREDITY ALGORITHM

ISFH algorithm is an evolutionary algorithm and it aims to obtain the best solution within a short time. This process is also applied in data mining for feature selection methods. In this paper, we are finding a solution to solve the classification problem using only improved sheep flock heredity algorithm. Improved sheep flock heredity algorithm is basically used for evaluating the natural evolution of sheep in a flock. ISFH algorithm simulates heredity of sheep flock in the lowland. Sheep in every flock are controlled by a shepherd. So that, the inheritance of the genetics can affect only the other sheep within the flock. Only, some special characteristics affect the sheep within the flock as well as in the nearest flocks. Those characteristics are

called as fitness characteristics which can breed in the flock. Two sheep flock may have mixed characteristics with other flocks. The better fitness characteristics can breed the most. The objective function considered in this paper is,

$$A_i \geq \delta \text{ and } A_i = \text{Trainingdata} - \text{value}, \forall i = 1, 2, 3, \dots, N.$$

ISFH ALGORITHM

1. Initialize and generate random population P
2. Make sub-chromosome using length x
3. Set the threshold values based on the training data for all elements in P called as OFV
4. Apply cross over on a sub-chromosome
5. Apply Inverse Mutation on sub-chromosomes
6. Compute OFV for mutated sub-chromosomes and compare with the parent chromosome OFV and choose the chromosome according to the best OFV
7. Apply Single Point Mutation on sub-chromosomes
8. Compute OFV for mutated sub-chromosomes and compare with the parent chromosome OFV and choose the chromosome according to the best OFV
9. Compute Robust Replace Heuristic [1/OFV]
10. Repeat the step 4 for chromosome levels until obtaining a best OFV

IMPROVED SHEEP FLOCK HEREDITY ALGORITHM PSEUDO CODE

Step-1: Read all the online transaction data with N number of attributes where $N = \{A_1, A_2, \dots, A_N\}$, using **IAgent**. Read all the Bank data with N number of attributes where $N = \{BA_1, BA_2, \dots, BA_N\}$, using **BAgent**. Read all the Credit Card data with N number of attributes where $N = \{CCA_1, CCA_2, \dots, CCA_N\}$, using **CCAgent**.

Step-2: The critical values are calculated by comparing the **Ai** values with the training data.

Step-3: For all the generations the A_i values are verified as Fraud or Normal using ISFH algorithm.

Step-4: Generate fraud transactions using ISFH algorithm and repeat step-1 to step-3. This process is used to analyze the feasibility of the CFD.

IMPROVED SHEEP FLOCK HEREDITY ALGORITHM

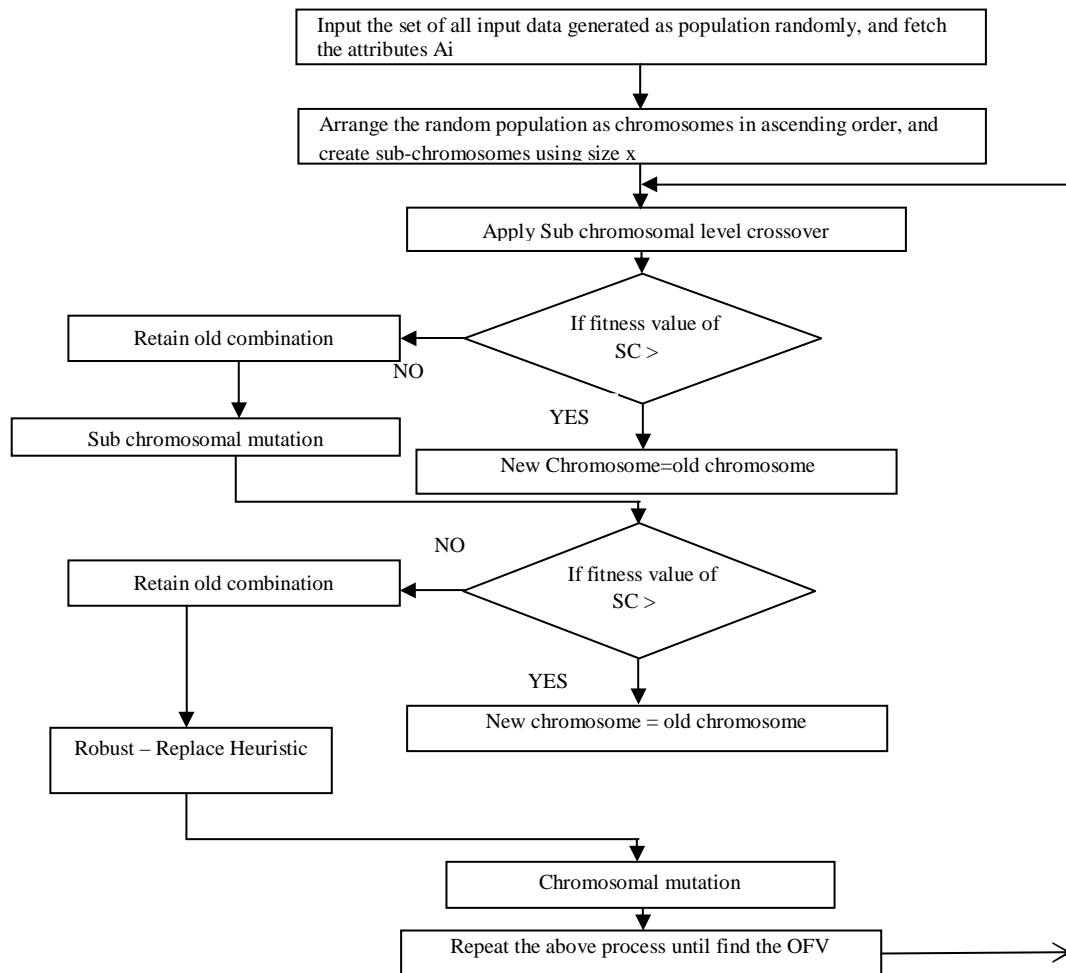


Figure-2: ISFH Algorithm Flow Diagram

IMPLEMENTATION

1. Initialize Population $P = \{ A_1 \text{ to } A_n \}$
 - a. Check the conditions $\left\{ \begin{array}{ll} \text{score} = i & \text{if } (A_i \text{ meets conditional value}) \\ \text{score} = 0 & \text{else} \end{array} \right\}$
 - b. According to the Credit Card Usage Frequency
 - c. $CCFreq = \text{Total Number of card used}(CU)/CCage$
 - d. If CC Freq is less than 0.5 then, If $(A1.valid = true) \&\& (A2 == Accept) \&\& (A3 \geq CC - amount) \&\& (A4 < BankBalance) \&\& (A5 \leq dueDate) \&\& (A6 \leq CardExpiry Date) \&\& (A7 \leq dueDate) \&\& (A9 \leq curTime+10s) \&\& (A15 == reg.IPaddress) \&\& (A16 \leq Bank of the card) \&\& (A5 == rec.history)$
 - e. then
 - a. $score = 1$
 - f. Else
 - b. $Score = 0$
2. Make Sub-chromosomes as $= \{ \{A_1 \text{ to } A_m\}, \{A_m \text{ to } A_k\}, \dots, \{A_l \text{ to } A_n\} \}$
3. Apply cross over and compute the critical score and compare it with the DB values and registered values
4. Apply inverse mutation and compute the critical score and compare it with the DB values and registered values
5. Apply the single point mutation and compute the critical score and compare it with the DB values and registered values
6. Repeat the above steps until all attributes satisfy all the conditions, then take that transaction as original transaction, and provide acceptance notification for further proceedings, else, reject the corresponding transaction as fraud.

IMPLEMENTATION RESULTS

To experiment and verify the performance of this proposed approach it is implemented in DOTNET software and four systems were used for that. One system is assumed as server, where the registration data are stored, the other system is treated as the bank and the bank database is stored, in the third system the credit card company data and card database is stored. Finally the fourth system is treated as user system from where the transaction is started. Since, all the systems are connected in Wi-Fi network, all the systems and software are distributed via DOTNET 2012 software, due to its interoperability.

Personal Details

Do you currently hold an SBI Credit Card ? : No Yes *Mandatory Fields

Name*: sumathi Loga Nathan

Date of Birth*: 16 May 1973 Gender*: Female Male

Qualification*: Post Graduate and Above PAN Number*: AB93837492

Email ID*: griou@gmail.com Mobile*: 9994048308

Res. Address Line 1*: Madurai Res. Address Line 2: Bungalow

City*: CHENNAI Pin Code*: 600052

Res. Number: 044 8746282 Occupation Type*: Self Employed

[SUBMIT LATER >](#) [NEXT >](#)

Please enter valid PAN no.

Figure-3: Constraints Verification in Credit Card Application Online

In online application, all the credentials are verified element by element. If the elements are valid, the next level of application can be permitted else, it gets rejected. Figure-3 shows that, the PAN number entered in the application form is not valid one, so it makes the customer to enter the valid PAN card number compulsory, else they should cancel themselves. This is the initial level of prevention applied for CFD.

In the offline application, all the credential values are fed into a database and each element values are fed into ISFH algorithm and verify all the attributes are best values or not. If the values are best values then that application form will be approved else it will get rejected. The Table-2 shows that some of the highlighted entries are fraud entries.

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	Score
22221	1	4232	4	10000	4	12000	12	5	0	0	0	0	1
22222	2	4342	45	4500	30	23000	23	6	9	2	0	9000	1
22223	3	5445	52	5300	5	12000	34	3	8	2	1	15000	1
22224	4	6453	64	5000	65	34000	45	7	7	14	0	85000	1
22225	5	6423	74	5000	23	9000	65	3	6	7	0	12000	1
22226	6	7634	34	6000	2	8000	54	3	3	0	1	12000	0
22227	7	9876	45	30000	5	120000	43	3	8	0	1	0	1
22228	8	1234	65	6500	7	10000	34	8	9	4	0	1900	1
22229	9	4325	76	54000	65	65000	32	54	8	2	0	16000	1
22230	10	5464	56	95000	65	100000	21	7	6	10	1	16000	1
22231	11	6536	34	10000	43	23000	11	4	7	2	0	11000	1
22232	12	7647	54	34000	5	45000	15	7	6	11	0	0	1
22233	13	9856	34	45000	43	56000	16	3	7	11	0	0	0
22234	14	3846	35	50000	7	60000	76	6	5	12	0	14000	1
22235	15	8473	67	4000	8	5000	87	3	4	3	0	9000	1
22236	16	3836	86	56000	12	129000	19	6	5	1	1	19000	1
22237	17	8476	34	55000	32	85000	87	3	3	12	0	0	0
22238	18	7363	67	45000	45	80000	86	6	5	3	0	19000	1
22239	19	5264	34	55000	67	65000	46	3	3	12	0	0	0
22240	20	7362	43	55000	89	89000	74	4	5	0	1	7000	1

Table-2: Experimented data

The entries 22226, 22233, 22237 and 22239 are having similar scores in attribute number 4, attribute number 10, attribute number 12 and similarity total score are rejected and it is shown in Table-2. From the overall dataset, a set of 100 samples are taken and analyzed using ISFH algorithm.

Total Entries	Normal	Fraud
100	23	77
99	22	77

Table-3: Fraud Detection

From the above recorded transactions, the transaction-ID 6, 13, 17 and 19 are the Fraud Entries due to much duplication in the entries. To evaluate the performance of the proposed approach, there are 100 entries were entered in the system and score is computed. From the 100 entries, there were 23 entries are detected as fraud, due to

the score. The remaining entries are suggested as normal entries and it is shown in Table-3 and in Figure-3.

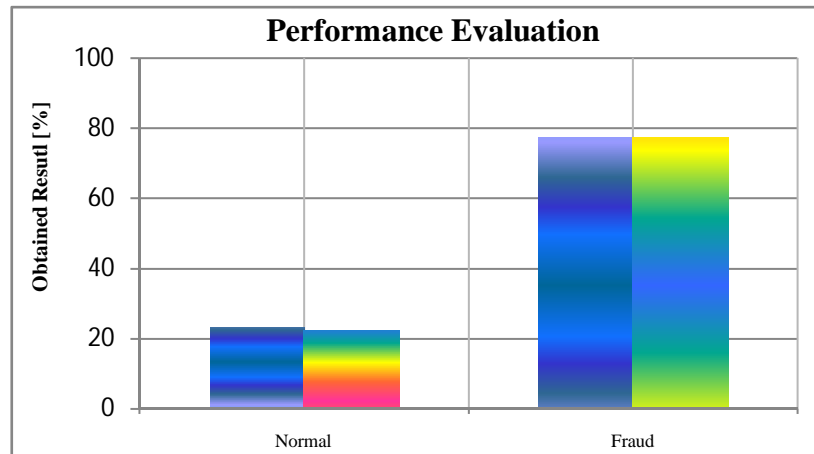


Figure-4: Fraud Detection

Our proposed approach detects and classifies the fraud data transaction 22 over 23 and classifies the normal transaction 77 over 77. The CFD system is needed in all the online transactions based on credit card, because, during the entry, the last entry is retained in the web browser cookie can be theft. So, it is necessary to clear the overall cookie and browser meta data after a successful transaction.

CONCLUSION

In this paper, the MLMA system with ISFH algorithm based optimization is presented for finding the credit card fraud detection and the results are examined in the applications of CC. The ISFH algorithm is used to execute the credit card fraud detection by verifying all the attribute values are in a particular range or not. In this paper, it is presented to detect the credit card fraud and the results are examined according to the principles of MLMA and ISFH algorithm. It can be seen that ISFH algorithm is applied to execute the CFD, to avoid fraud occurring in a financial institution to a customer or merchant. ISFH algorithm is a heuristic algorithm, which is used in this paper to obtain a better optimal solution. All the fraud data are detected by the ISFH algorithm and providing prevention for CFD. From the results and Figures, it is clear that the proposed approach is better than the other approaches can provide prevention for CFD also saves the time and un-wanted fund transfer.

REFERENCES

- [1]. Mike Ebbers, “5 Things to Know About Detecting Credit Card Fraud” – IBM-RedBook.
- [2]. Sam Maes, Karl Tuyls, Bram Van schoenwinkel, Bernard Manderick. Credit Card Fraud Detection Using Bayesian and Neural Networks First International NAISO Congresson Neuro Fuzzy Technologies, Havana, Cuba. 2002.
- [3]. M.J. Kim and T.S. Kim, “A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection, ”Proc. International Conference on Intelligent Data Engineering and Automated Learning, Lecture Notes in Computer Science, Springer Verlag, no. 2412, pp. 378-383, 2002.
- [4]. Dr.R.Dhanapal, “An intelligent information retrieval agent”, Elsevier International Journal on Knowledge Based Systems 2008.
- [5]. Binu Thomas and Raju, “A Novel Fuzzy Clustering Method for Outlier Detection in Data Mining”, International Journal of Recent Trends in Engineering, Vol.1, No.2, May 2009.
- [6]. Foster, D. & Stine, R., 2004. ‘Variable Selection in Data Mining: Building a Predictive Model for Bankruptcy’ .Journal of American Statistical Association, 99; 303-313.
- [7]. Pago-Report. 2005. The development of E-commerce sectors, ©Pago eTransaction Services GmbH.
- [8]. SHAIRESH S. DHOK, “Credit Card Fraud Detection Using Hidden Markov Model”, IJSCE-ISSN: 2231-2307, Volume-2, Issue-1, March 2012.
- [9]. Ghosh, S., and Reilly, D.L., 1994. Credit Card Fraud Detection with a Neural-Network, 27th Hawaii International I Conference on Information Systems, vol. 3 (2003), pp. 621- 630.
- [10]. Molyneaux, D. 2007. ‘Two case study scenarios in banking: a commentary on The Hutton Prize for Professional Ethics, 2004 and 2005’. Business Ethics: A European Review, 16:4, 372-386.
- [11]. Anderson, R. 2007. The Credit Scoring Toolkit: theory and practice for retail credit risk management and decision automation. New York: Oxford University Press.
- [12]. Linda Delamaire, Hussein Abdou, and John Pointon, “Credit card fraud and detection techniques: a review”, Linda Delamaire-Banks and Bank Systems, Volume 4, Issue 2, 2009.
- [13]. Euro monitor International, 2006. Financial cards in Germany Available at: http://www.euromonitor.com/Financial_Cards_in_Germany (Accessed: November 2006).
- [14]. FSTC :<http://www.fstc.org/>
- [15]. K.RamaKalyani, D.UmaDevi, “Fraud Detection of Credit Card Payment System by Genetic Algorithm”, -IJSER-Volume 3, Issue 7, July-2012.
- [16]. <https://www.maxmind.com/en/minfraud-services>.

Author's Biography

Mrs.V.Mareeswari, is pursuing her PhD degree in Data Mining at St.Peter's University of Computer Science Engineering, Avadi ,Chennai ,Tamilnadu, India. She received her M.E degree in computer science and engineering from Madha Engineering College, Kunrathur, Chennai, Tamilnadu, India, in 2011. She received her B.E degree in Electrical and Electronics Engineering from GCT Campus, Anna University, Coimbatore, Tamilnadu ,India ,in 2003. Her interests include Data Mining and Image Processing.



Dr.G.Gunasekaran is completed his PhD degree in Data Mining at Jadavpur University, Kolkata ,India in 2009. He received his M.E degree in computer science and engineering from Jadavpur University, Kolkata, India, in 2001. He received his B.E degree in Computer Science and Engineering from Madurai Kamarajar University, Madurai, Tamilnadu, India, in 1989. His interests include Data Mining, Bio Informatics, Software Engineering, Graphics & Multimedia.