Scientific Journal Impact Factor (SJIF): 1.711
e-ISSN: 2349-9745
p-ISSN: 2393-8161
International Journal of Modern Trends in Engineering and Research
www.ijmter.com

CREDIT CARD APPLICATION FRAUD DETECTION SYSTEM USING MULTILAYER ALGORITHM

Mr.Vipul Agrawal [1], Prof. Vijay Kumar Lokhande[2]
[1] *PG Student Department of CSE, BIST,Bhopal*
[2] *Head Department of CSE, BIST,Bhopal*

Abstract— Identity crime is well known, prevalent, and costly; and credit application fraud is a specific case of identity crime, involving synthetic identity fraud and real identity theft. Identity crime has become prominent because there is so much real identity data available on the Web, and confidential data accessible through unsecured mailboxes. It has also become easy for perpetrators to hide their true identities. Credit card fraud is an element of identity fraud. It can have far reaching effects, since the information on the card can be used to perpetrate other types of identity theft crimes. The existing non-data mining detection system of business rules and scorecards, and known fraud matching have limitations.

To address these limitations and combat identity crime in real time, this system designs a new multilayered detection system complemented with two additional layers: communal detection (CD) and spike detection (SD).CD finds real social relationships to reduce the suspicion score, and is tamper resistant to synthetic social relationships. It is the white list-oriented approach on a fixed set of attributes. SD finds spikes in duplicates to increase the suspicion score, and is probe-resistant for attributes. It is the attribute-oriented approach on a variable-size set of attributes. Together, CD and SD can detect more types of attacks, better account for changing legal behavior, and remove the redundant attributes.

Although this research is specific to credit application fraud detection, the concept of resilience, together with adaptivity and quality data discussed in the paper [1], are general to the design, implementation, and evaluation of all detection systems.

Keywords- Communal Detection; Data Mining; Data mining-based fraud detection ;Identity crime; Spike Detection. Identity crime, Multilayer Algorithms et. al.

I. INTRODUCTION

As Fraud Detection is imperative for every business or organization as it impacts by increasing the cost of doing their businesses. Identity theft is a form of stealing someone's identity in which someone pretends to be someone else by assuming that person's identity, usually as a method to gain access to resources or obtain credit and other benefits in that person's name. The victim of identity theft (here meaning the person whose identity has been assumed by the identity thief) can suffer adverse consequences if they are held responsible for the perpetrator's actions. Identity theft occurs when someone uses another's personally identifying information like their name, identifying number, or credit card number, without their permission, to commit fraud or other crimes. Credit application fraud is a specific case of identity crime [1]. Identity crime has become a more common approach as there is so much real identity data available on the Web, and confidential data accessible through unsecured mailboxes. It has also become easy for fraudster to hide their true identities. Customers fill the credit application form online or using manual paper based work.

When online forms are not closed properly, or when the paper forms are not filled or discarded safely, it becomes easy for the fraudsters to misuse the customer information. It leads to the level of deciding the economy of the country.

Most of the businesses implemented intelligent analytical techniques to control fraudulent

activities to reduce their increased cost. One of the intelligent techniques called data mining, substantiated to be a competitive tool for controlling fraudulent activities. Data Mining can extract valuable patterns or knowledge from large volumes of data and alarm fraud. Data mining can be defined as the process of finding previously unknown patterns and trends in databases and using that information to build predictive models. Alternatively, it can be defined as the process of data selection and exploration and building models using vast data stores to uncover previously unknown patterns [3]. Data mining is not new; it has been used intensively and extensively by financial institutions, for credit scoring and fraud detection. The data mining consists of multiple algorithms for detection [5]. Data mining algorithms are used in the online credit card application for counterfeit detection.

The main objective of this paper is to highlight the use of efficient data mining algorithms in credit application fraud detection system. The algorithms are used in this system is the spike detection and communal detection together.

II. LITERATURE SURVEY

2005, Efstathios Kirkos et al. explores the effectiveness of Data Mining (DM) classification techniques in detecting firms that issue fraudulent financial statements (FFS) and deals with the identification of factors associated to FFS.

2009, G. Apparao et al. analyzes that the prevention is the best way to reduce frauds, fraudsters are adaptive and will usually find ways to circumvent such measures.

2011, Tatsuya Minegishi et al. focus on classification learning, which is an analytical method of stream mining.Peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

2012, Sherly K.K et al. evaluates three classification methods to solve the fraud detection problems for data mining and shows how advanced techniques can be combined successfully to obtain high fraud coverage with maximum confidence and minimum false alarm rate. In 2012, Clifton Phua et al. observe that the credit application fraud is a specific case of identity crime. The existing non-data mining detection system of business rules and scorecards, and known fraud matching have limitations. To address these limitations and combat identity crime in real time, they propose a new multilayered detection system complemented with two additional layers: communal detection (CD) and spike detection (SD).

III. PROS AND CONS

Due to the extensive growth of E-Commerce fraud detection has become a necessity. Fraud detection is a continuously evolving discipline and ever changing tactics to commit fraud. In Real Time the Non-Data mining approach of business rules, scorecards and known fraud matching technique is used for the crime detection on credit cards which has some limitations. In case of the existing system the fraud is detected only after the fraud is done. To overcome these drawbacks the System uses data mining techniques to prohibit the crime on credit cards before it has been done. This method provides a way to prevent fraudsters from receiving credit cards, rather than finding way to detect fraudster after the card has been issued. Data mining techniques selected for the fraud detection was based on the effectiveness of the adaptive economic environment. The system concentrated on improving the fraud detection systems accuracy, time and cost.

IV. IMPLEMENTATION DETAILS

The system utilizes two layers called Communal Detection (CD) and Spike Detection (SD). The main contribution of this paper is to enhance secure transaction in credit card applications by using two new data-mining layers. These new layers improve detection of fraudulent applications because the detection system can detect various kinds of attacks, better account for changing legal behavior, and eliminate the redundant attributes.

4.1 Communal Detection

Communal Detection layer is based on white list-oriented approach. It utilizes fixed set of attributes. Communal Detection (CD) finds real social relationships to reduce the suspicion score, and is tamper- resistant to synthetic social relationships [1]. The CD algorithm matches all links against the white list to find communal relationships and reduce their link score. CD has a fundamental weakness in its attribute threshold. Specifically, CD must match at least three values for our data set. With less than three matched values, our white list does not contain real social relationships because some values, such as given name and unit number, are not unique identifiers. The fraudster can duplicate one or two important values which CD cannot detect.

With this data stream perspective in mind, the CD algorithm matches the current application against a moving window of previous applications. It accounts for attribute weights which reflect the degree of importance in attributes. The CD algorithm matches all links against the white list to find communal relationships and reduce their link score. It then calculates the current application's score using every link score and previous application score. At the end of the current micro discrete data stream, the CD algorithm determines the SoA and updates one random parameter's value such that it trades off effectiveness with efficiency, or vice versa. At the end of the current Mini discrete data stream, it constructs the new white list.

4.1.1 CD Algorithm–

1. Every application value is compared against a list of previous application values to find the links.
2. Every application's link is matched against the white list to find communal relationships among applications and reduce their Link score.
3. Every previous application's score is to be included into the current application's score. Previous score acts as a baseline level.
4. Calculate every current application score using link and previous application's score.
5. The algorithm updates one random parameter's value such that there is a tradeoff between effectiveness with efficiency, or vice versa.
6. A new white list is constructed on the current Mini-discrete stream links.

4.2 Spike Detection

In contrast to CD, SD finds spikes to increase the suspicion score, and is probe resistant for attributes. Probe resistance reduces the chances a fraudster will discover attributes used in the SD score calculation. It is the attribute- oriented approach on a variable-size set of attributes. SD complements CD. The redundant attributes are either too sparse where no patterns can be detected, or too dense where no denser values can be found. The redundant attributes are continually filtered; only selected attributes in the form of not-too-sparse and not-too-dense attributes are used for the SD suspicion score. In this way, the exposure of the detection system to probing of attributes is reduced because only one or two attributes are adaptively selected [1].

From the data stream point-of-view, using a series of window steps, the SD algorithm matches the current application's value against a moving window of previous application's values. It calculates the current value's score by integrating all steps to find spikes. Then, it calculates the current application's score using all values scores and attribute weights. Also, at the end of the current Mini-discrete data stream, the SD algorithm selects the attributes for the SD suspicion score, and updates the attribute weights for CD.

4.2.1 SD Algorithm

1. Every application value is compared against a list of previous application values step by step.
2. Calculate application's current value score by integrating the steps to find the spikes.
3. Calculate application's score using attribute weights.
4. Identify the key attributes to calculate the SD suspicion score.
5. The final step updates the weights of the attributes.

V. SYSTEM ARCHITECTURE

The figure 1 illustrates the system architecture-

The Description of System Architecture is given below [2] -

[1] Display GUI for entering credit card application details.

[2] Take input from user.

[3] Compare new application with each other application in the bank database and assign a link type. The link type is nothing but a binary string (e.g. 01000101) in which "1" represents matched fields and "0" represents unmatched fields.

[4] Initial white list is created. The White list has list of verified applications, link type, number of applications corresponding to a particular link type and weight.

[5] Communal Detection -

a) New Application is compared with windows of applications in the white list.
b) CD layer finds communal relationships between the applications.
c) If four or more fields are matched in the new application against application in the white list, then CD assigns less suspicious score.
d) Otherwise the new application form is added into the white list and the list is updated.
e) Suspicious score assigned to new application form is given as input to the SD layer.

[6] Spike Detection –

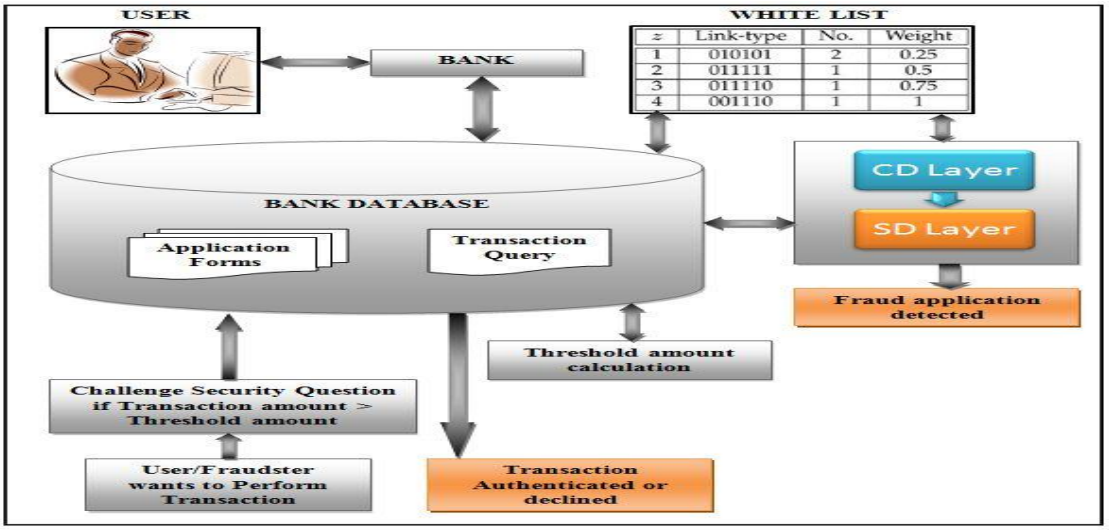
a) Spike Detection (SD) layer verifies the matched fields for their priority. The unique ID fields are given higher priority.
b) If unique IDs are matched then the suspicious score gets increased and the application form is declared as fraud and hence finally rejected.
c) If none of the unique IDs are matched then the application form is added into the white list and the list is updated.

[7] Threshold Transaction Amount Calculation –

Based on the previous transactions made by the user the system calculates a threshold value of the transaction amount. The threshold value is nothing but average of all the previous transactions.

[8] Secure Transaction -

Now the fraudster or the legal user performs credit transaction. If the credit transaction amount is higher than the threshold, the fraudster or legal user is asked to challenge the security question. If the challenge is success i.e. in case of legal user the transaction is authenticated otherwise it is declined in case of fraudster. Hence the secure transaction is performed.



z	Link-type	No.	Weight
1	010101	2	0.25
2	011111	1	0.5
3	011110	1	0.75
4	001110	1	1

Fig. System Architecture

Fig.1: System Architecture

VI. RESULT DISCUSSION

The input data set for the system is a synthetic data set of 50,000 credit applications which is available at https://sites.google.com/site/cliftonphua/communal-fraudscoring-data.zip.There are about 30 attributes are present in synthetic data set out of which only 19 most important identity attributes are selected for the processing. The system will take input credit application for credit card from the user through online web portal and the application submitted by user act as a real-time entry to the system.

The system will compare the application submitted by the user with applications in the synthetic data set and validate the application. The white-list is constructed form the input data set and a CD suspicious score is assigned to each application as a result of communal detection algorithm. The figure 5 shows the sample white-list constructed from credit applications in figure 2 and figure 3.

The Spike detection algorithm outputs the SD suspicious score. CD and SD scores are combined together to give a single score. SD updates the CD attribute weights.

This dataset is set of accepted list of credit card applications. We consider here 5 records are in accepted list with 12 attributes and with 2 unique attributes adhar card no and PAN id with status accepted. Record 6 and 7 are the input dataset to the system. Here the status of the records is in pending status the figure 3 shows the White-List constructed by the CD algorithm as a RESULT DURING PROCESSING THE INPUT DATASET.

When we apply CD and SD algorithm on it calculates the link-type, and weight of the attributes. And after successful execution the status of the record is changes from pending to accepted or rejected.

Rec_id	First Name	Last Name	Address	City	State	Postcode	MobileNo	Adhar No	PAN_id	DOB	Status
1	kale	noyce	bural court	forbes	vic	423075	7532950055	123456789101	AMZGT1000A	11/10/1947	Accepted
2	bella	chemny	howie court	nambour	vic	422540	7558336354	123456789102	BNAHU1001B	29/01/1931	Accepted
3	charlotte	bullock	bungaree crescent	wagoora	qld	422154	7570610247	123456789103	COBIV1002C	10/10/1982	Accepted
4	esme	jardine	kirby place	woodville north	nsw	423233	7559190277	123456789104	DPCJW1003D	30/10/1949	Accepted
5	andrea	rothe	kinsella street	kumbia	nsw	424305	7532379831	123456789105	EQDKX1004E	19/06/1922	Accepted

Fig 2: Dataset in Accepted List

6	thomas	matthews	maranoa street	marayong	nsw	423104	7573884029	123456789107	GSFMZ1006G	27/12/1932	Pending
7	anurag	sangale	bungaree crescent	wagoora	qld	422154	8806323532	123456789103	ABCDE1234F	22/12/1990	Pending

Fig 3: Input data set

Link-type	Count	Weight
0000100101	1	0.11
0000010101	1	0.22
0000010101	1	0.33
0000010101	1	0.44
0000000101	1	0.55
0000000101	1	0.66
0000100101	1	0.77
0000110101	1	0.88
0011010100	1	1

Fig 4: link type

6	thomas	matthews	maranoa street	marayong	nsw	423104	7573884029	123456789107	GSFMZ1006G	27/12/1932	Accepted
7	anurag	sangale	bungaree crescent	wagoora	qld	422154	8806323532	123456789103	ABCDE1234F	22/12/1990	Rejected

Fig 5: Output Dataset

VII. CONCLUSION

Main focus of this project is the detection of fraudsters in credit applications by implementing a new multilayered detection system complemented with the new data mining layers Communal Detection(CD) and Spike Detection(SD) together which helps in performing a secure transaction in real-time. The implementation of CD and SD layers is done to detect fraudulent activities in duplicates as well as the real social relationships. Communal Detection and Spike Detection layers are continuously updated so that the fraudster should never get a chance of attacking again. It has documented the development and evaluation in the data mining layers of defense for a real-time credit application fraud detection system. The main focus of this project is the real-time search for patterns in a principled fashion, to safeguard credit applications at the transaction. The system

implements the concepts of resilience (multilayer defense), adaptivity (accounts for changing fraud and legal behavior), and quality data (real-time removal of data errors.

REFERENCES

[1] Clifton Phua, Kate Smith-Miles, Vincent Cheng-Siong Lee and Ross Gayler, "Resilient Identity Crime Detection", IEEE Transactions on Knowledge and Data Engineering, vol.2, no. 3,pp.533-546, 2012.
[2] Alka Herenj, Susmita Mishra, "Secure Mechanism for Credit Card Transaction Fraud Detection System", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 2, February 2013.
[3] Namrata Shukla, Shweta Pandey, "Document Fraud Detection with the help of Data Mining and Secure Substitution Method with Frequency Analysis", International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume 2, Number 2, June 2012.
[4] K. Vidhya, P. Dinesh Kumar, "Multi-Secure Approach for Credit Card Application Validation", International Journal of Computer Trends and Technology, volume 4,Issue 2, 2013.
[5] M. Swathi, K. Kalpana, "Spirit of Identity Fraud And Counterfeit Detection", International Journal of Computer Trends and Technology (IJCTT) , volume 4, Issue 6, June 2013.
[6] Clifton Phua, Kate Smith-Miles, Vincent Lee and Ross Gayler- Adaptive Spike Detection for Resilient Data Stream Mining, 2010.
[7] T. P. Latchoumi, V. M. Vijay Kannan, "Synthetic Identity of Crime Detection", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 7, July 2013.

