ISSN:0975-9646

Algorithm Based Security Infrastructure in Online Fraud Detection.

Rupal Harishchandra Jadhav, Swapnali Sharad Hinge, Shahabaj Sanaullah Shaikh, Swapnil Subhash Kanade, Prof. K.S.Wagh.

Department of I.T. UoP,
MMIT,Pune-411043, Maharashtra, India

***Abstract*-In this paper, some techniques are use for online fraud detection such techniques as data mining and data model with naive bayes possibility. In that mining we use specific areas like time, amount, area when any area get varied then system get alert and start verification. After verification process result system decides to lead transaction or stop transaction. We are showing this concept with example.**

Keyword- fraud identification; data mining; Naive Bayes possibility.

I. INTRODUCTION

Finance is the biggest concept in any field and crime happens according to field and money. For that many technology's are developed and form with the help of these to prevent fraud but, not stop as well as not give guarantee about fraud happen in next time. In any financial organization store large information about capital and there details [1].

Data mining is very popular as well as strong concept in real life. We refer data mining as data mining is nothing but the mining of knowledge from data. It is technology of extracting useful information from big databases. Data mining is deals with previous unknown, valid, hidden information for making the decisions. The interaction between actions, behaviours and detection of abnormal condition we obtain from various areas observation [2]. Also data mining is useful for transformation of raw data into necessary information from data. Data mining analyze data to find useful information from data set. Data mining is process with part prediction, classification and clustering.

A. Prediction: The data that we need for decision making that requires prediction. e.g. credit card companies predict person with authentication for security. In our system, prediction is based on Amount, Area & Time. Rules for making the prediction are derived from the same attributes of past & current observed behaviour. We can determine, feature data states based on past & current data.

B. Classification: In classification, predefined groups or classes are made by mapping the data. Classification is done by finding the rules that partition the given data into different groups. Example Credit Card Company decides whether or not to give a credit card by classifying. Classification finds the new records by classifying from the root to the leaf through the branches.

C. Clustering: Clustering is defined as a grouping of data but groups are not predefined. The partitioning and segmenting the data into groups that might or might not be different. Clustering is a process of partitioning the data set into cub classes. Clustering is usually accomplished by determining the similarity among the data on predefined attributes. The most similar data are group into cluster. Clustering is useful for users to understand structure from the data set.

II. CHALLENGES

- To handle large amount of data with proper inputs from end user.
- There are various transactions according to specified areas at that time the calculation of possibility is nearest.
- Transactions are not straight-way there are various variations as per users.
- The integration of data sets is only in useful form.

In paper, we are clear all challenges as four ways initial way we generate database with simple form according to the attributes and its values. Second the transaction are in thousands of number these possibilities handled with integration of one possibility to another possibility then it gives result as more strong ways as well as more reliably. For that calculate with current transaction and without current transaction and compare. Third the varied transaction input can handle with specific format access of data such that give amount ranges for transactions. The integration of data sets are useful by timing part integrate with areas part those are integrate with amount part [3].

III. ESTABLISHMENT

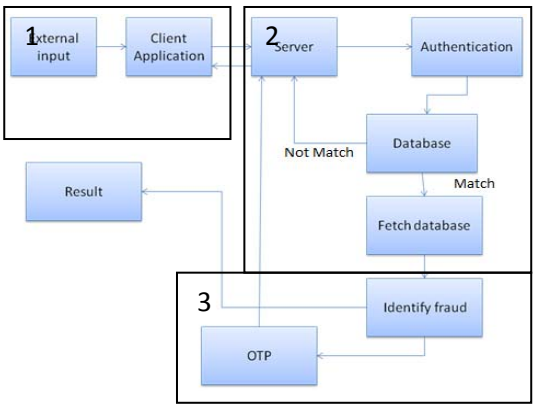

Fig. 1. Model of online fraud detection.

In this system having 3 main part as show in fig.1. 1: External input, 2: Server with database 3: Identification. These parts show with rectangle in block diagram.

A. *External Input:* External input is nothing but end user request is in the form of amount, time and area. These external input support most important data about your account detail.

B. *Server Database:*
Server serves services to the client same way server serve unauthorized identification of person who access your details. This identification happen by data mining technique with Naive Bayes algorithm. This algorithm leads with help of areas such as amount, time, areas.

C. *Identification:*
OTP(One Time Password) is more stronger technique to identification of actual unauthorized person by sending SMS OTP to authorized person if it is verifies correctly then transaction move a head otherwise, it will stop.

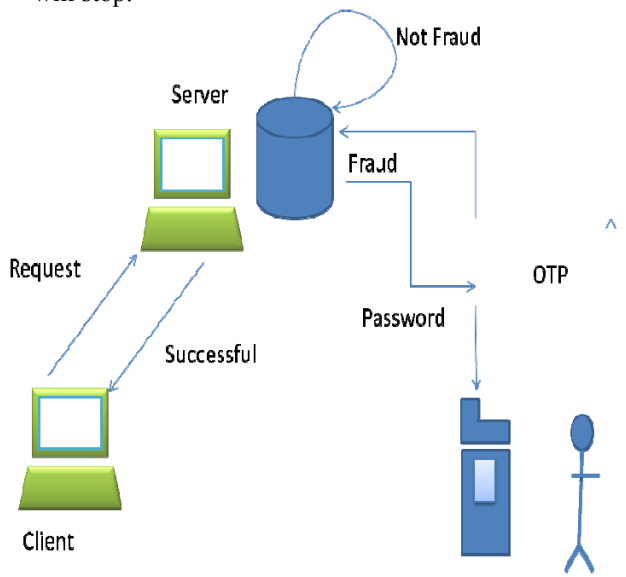

Fig. 2. Actual application of detection.

Figure show actual implementation about online fraud detection application.

First end user send or to give requirement to server using these requirement server get identified that user is valid or not with data mining support in which there are three areas are consider first amount, second area and third duration when any area get varied according to attribute values, if user is valid then transaction move next and all details are update in data base with serialization-deserialization else system alert and move toward the authentication. In this part systems generate OTP that means unique password at every time for valid end user if valid user insert valid password then system move to normal state and successfully done work otherwise authentication process follow until valid user not found.

Bayes theorem having bayes rules, this rules consists condition based probability and joint based probability. But condition based probability is more basic than joint based probability.

$P(F/T) = [P(T/F)P(F)] / P(T)$

Where,

F = For transaction in fraud.

T = For time accordingly transaction.

Naive bayes algorithm is totally based on bayes theorem. Bayes theorem is follow probability estimation and estimation is follow trained dataset this trained dataset is nothing but already presented data in database. The naive bayes algorithm is meagerly supported to data mining applications by calculation of possibility.

The same rule we use for many area. When any one variation found in possibility then system detect that fraud happen. The number of possibilities area increases the result is security also increases.

Main advantage of naive bayes is that it requires trained dataset for decision making this decision is most of right form for what happen in future. Also, when performance of this calculation does not affect on actual data in database.

Naive Bayes also handles in distribution database by accessing sensitive data from database. Naive Bayes work independently that means work specifically area or attribute values. This working method of naive bayes, helpful for analysis of probability of possibility.

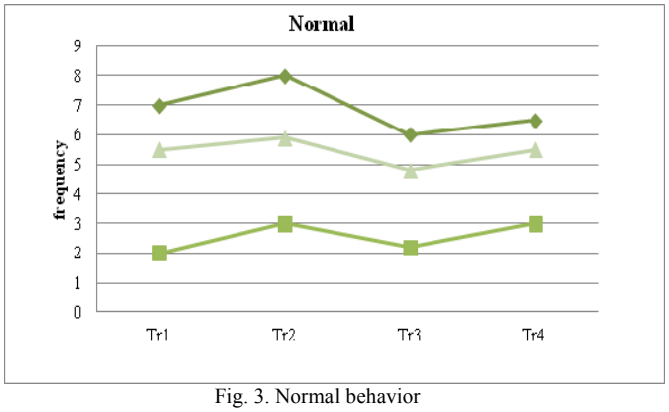

Fig. 3. Normal behavior

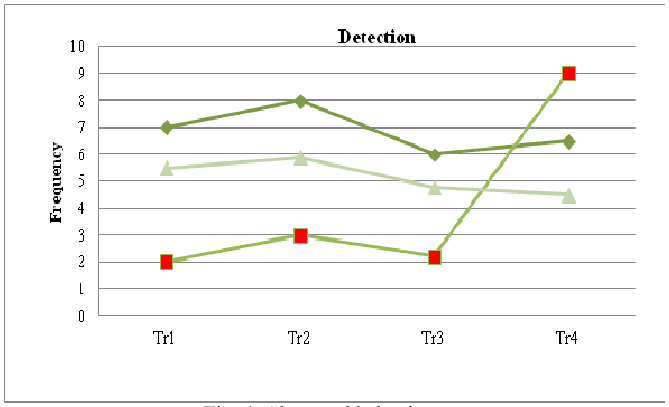

Fig. 4. Abnormal behavior

Fig. 3. System at normal condition. Fig. B. System at abnormal behavior [3]. Fig 3. shows normal data accept without any authentication and update database by append external data in existent database.

Fig. 4. show that when any abnormal input come then system get alert for transaction and verifies if verification is successful then system update its own database.

TABLE 1: Duplicate database with some attributes and its values.

Id.	Type	Time	Date	Day	Area	Amount
1.	Withdrawal	10:00	1	Monday	EMI	200.00
2.	Withdrawal	11:10	2	Tuesday	EMI	250.00
3.	Deposit	11:50	2	Tuesday	-	10000.00
4.	Withdrawal	12:02	3	Wednesday	Car Maintains	700.00
5.	Withdrawal	12:20	4	Thursday	Medical	800.00
6.	Withdrawal	15:10	5	Friday	Fees	500.00
7.	Withdrawal	16:07	6	Saturday	Movie	200.00
8.	Withdrawal	11:00	6	Saturday	Recharge	1000.00
9.	Withdrawal	12:00	7	Sunday	Clothes	7000.00
10.	Withdrawal	10:30	8	Monday	EMI	210.00
Current Input	Withdrawal	10:29	8	Monday	Clothes	10000.00

TABLE 2: Duplicate database with amount attribute and its values rages.

100-500	501-1000	1001-1500	1501-2000	…….	6500-7000
5/9	3/9	0/9	1/9	……	0/9

TABLE 3: Duplicate database with area attributes and its values.

EMI	Car Maintains	Medical	Fees	Movie	…….	Clothes
3/9	1/9	1/9	1/9	1/9	……	1/9

IV. NAIVE BAYES IN PRTDICTION METHOD.

P(F/T) = [P(T/F)P(F)] / P(T)

Calculate probability without current transaction.

P(F/T) = [P(T/F)P(F)] / P(T)

Calculate probability with current transaction.

And then after compare if difference is large then system get alert as illustrated in table 1, table 2 and table 3..

In our system inputs are come with specified areas at that time any area get varied system get identified with help of data mining model as well as Naives Bayes Algorithm as.

2/9*2/2*2/2*2/2=0.2223

This value comes on existing database same ways:

3/10*3/3*1/3*1/3

enerate from current inputs 0.03 these value is less than existing data analysis at that time system get alert and start verification when system verified that user not authorized then transaction not move ahead.

Ai (∑ Specific transactions / ∑ transactions before current)

Pj = Ai-N-1 * Ai-N-2 *…… Ai-N

Where,

Ai : Specified areas according to system.

Pj: Probabilities that comes inour system

Same formula use for calculate both Probability that comes in our system at that time system consider minimum value because minimum probability is act as quantity of transaction that means possibility

V. CONCLUSION

In this paper, online transaction fraud detection is explain clearly and establish strongest application these is depend on data mining with specific areas, these areas give more effective view to tack decision as fraud occur or not. When occur any fraud then system handle more sophisticated way handle these fraud. It play effective role after developing possibility of areas. This concept we can handle on any other sensitive data like telecommunication, medical and scientifically developed information.

REFERENCES

[1] Hongzhi Yu Fengxin Liu Kaiqi Zou, Wenming Sun. Id3 decision tree in fraud detection application. IEEE International Conference on Computer and Electronics Engineering, 2, 2012.

[2] Member IEEE Pablo Samuel Chao Chen, Daqing Zhang. iboat: Isolation-based online anomalous trajectory detection. IEEE TRANSACTION ON INTELLIGENT TRANSACTION SYSTEM, 14.

[3] Clifton Phun. Resilient identity crime detection. IEEE TRANSACTION ON KNOWLEDGE AND DATAENGINEERING, 24.