Internet Banking Security Enhancement Using Naïve Bayes Algorithm

Ganesh Kumar, Achintya Ranjan Chaudhary, Kishan Kumar

***Abstract-** This paper explains the detail implementation of online banking or e banking authentication system. Security is a very important issue for online banking application which are being implemented by various online platform that are making the world getting closer. When implementing online e banking system, secure money transfer is very important which can be achieved by using database encryption and time based OTP. Techniques for secure storage of sensitive data are also used to prevent any intrusion. To reduce threat of phishing and to confirm user identity and location which would be known by coordinates generated by user's mobile device can be used as a method to prevent the weakness of traditional password based system can be improved by one time password which is sent if there is a major deviation in the behavior of the user. Challenge-based like asking less personal questions and time-based OTPs which is generated on mobile provide strong security because their period of validity is controlled directly by the bank and does not depend upon user's behavior. It is advised that the banks should not allow the OTP time window to exceed 5 minutes for a OTP. Based on the trust factor, the purpose of this paper is to examine the role of trustworthiness and trust in users' intentions on a transaction to be made and discontinuing the anomalous users from completing the transaction.*

***Keywords:** Encryption, Trustworthiness, Phishing, One Time Password.*

I. INTRODUCTION

Electronic trade is the fundamental accomplishment of the data and correspondences innovation in different financial fields. This innovation has added to business advancement, encouraged correspondence between financial components, prepared for the activity of little also, medium endeavours, advanced efficiency, and spared time and cash. data and correspondences innovation has expanded similarity among organizations and prompted the making of new openings for work. Opening of new accounts and transacting money across accounts, online shopping and transfer etc has been increased enormously these days. The accessibility of the web to a huge number of clients together with the facility of a large online platform and money portal have make clients works very easy for the money related works. Now every clients have just a user id and a password to login and spend, but what if your password is stolen or got phished or hacked.

The hacker just need one big transaction to steal all your money in one go, by the time clients know their bank balance, it will fell down to null, leaving the customers shattered. What if we have a system that knows me well enough that if I don't transact it will detect a anomaly. For this the system is first trained by the customer for the first 10 transactions. After that the system will create a threshold values in various parameters and will surely detect any suspicious transactions, next time. If the user is genuine and he is actually making that transaction there shouldn't be any kind of problem, but if the user is suspicious and the system detects it the user have to pass through challenge-based and time-based OTPs. The bank will send an otp to the suspicious user and ask him to verify failing bank may block the user. This will make the bank and the user more secured and protected. And will give a pause on the penetration and fraud in the field of e-banking. The bank will then obviously gain the trust of the user and will be very popular in its competitive field and will flourish in its sector for long.

II. LITERATURE SURVEY

Many institutes across the world have found suspicious activities in their financial systems. To deal with these issues, various methods for detecting suspicious activities have been encouraged, including checking of activities of employee on daily basis, law enforcement questions, and scanning clients record. Therefore, a system is needed that can expertly detect suspicious behavior and transactions and can take necessary steps based on previous pattern and algorithm. Previously researchers used a system that incorporates various fields of knowledge like electronic commerce, investment in integration software, information integration, database tracking expert information systems, knowledge management and many more. Recently, several other methods for detecting these suspicious transactions have been developed based on the machine learning algorithms, including dynamic Bayesian algorithm combined with fuzzy networks. Many ways and method have been brought day by day to stop the fraudulent behavior and activities in internet banking sector, some of them are discussed here. The bank that provides the services of e banking have tested various methods to identify the crime and track the customers behavior. Methods that are currently in use are transaction tracking, verification, personal identification number and finger prints. Biometric are also used that combinedly includes fingerprint and signature for the verification process of customers. Then keeping a record like a log book of good customers and bad customers and categorizing them according to different parameters like amount place etc.

Revised Manuscript Received on May 05, 2019.

Ganesh Kumar, Associate Professor, SRM Institute of Science & Technology, Kattankulathur, Chennai

Achintya Ranjan Chaudhary, SRM Institute of Science & Technology, Kattankulathur, Chennai

Kishan Kumar, SRM Institute of Science & Technology, Kattankulathur, Chennai

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Various Data-mining algorithms are used for this. Data Mining algorithm are based on specific learning pattern and rules that clearly be able identify any fraud activities. These are suitable and can handle a large amount of database. These algorithms and rules are used mainly for the detection of suspicious and fraudulent customers. The result of all these are used to create a alarm and try to stop the transaction or revert it. Association rule was then considered as good option for these results. Association rule aims at finding a set of data that enables us to predict the occurrences of other set of data from a given set of transactions. This method was basically used for extracting the credit card data, and to find unconventional or some suspicious pattern in the transaction, and furthermore identifying the user blocking it and preventing the fraudulent. The other method that was common was the Artificial neural network; it was also very useful in detecting and preventing the forgery. This method was capable of finding patterns of the customers that they leave in their previous transactions. Artificial neural networks can be trained with more different and vibrant data, so they are smart enough to adapt to new form of crime and fraudulent. This paper mainly aims at finding the suspicious behaviors in the internet banking by using Naïve Bayes Algorithm.

Mahdi Zamani and Mahnush Movahedi were reviewed many important algorithms for intrusion recognition supported by various Machine Learning techniques. Characteristics of Machine Learning techniques makes it potential to style ID'S that have very high detection value and low negative rates whereas the system quickly adapts itself to new upcoming frauds and intrusion. The algorithm is divided into 2 classes: 1. Artificial Intelligence

2. Computational Intelligence. Although these 2 have many similarities, many options of CI techniques, like fault tolerance, adaptation, high speed and error free even for the noisy data, adjusting the need of intrusion detection systems economically. Although a lot of range of techniques are used for finding the intruders but they maybe slow or unproductive. If clustering is used, the intrusion detection rate gets less.

R. Shanmugavadivu and Dr. N. Nagarajan have developed an anomaly mainly based on intrusion detection system to detect the fraud behavior in the computer network. A fuzzy based decision-making dataset was developed to build the system with a lot of accuracy for attack detection, using the fuzzy logic approach. An efficient set of fuzzy rules for reasoning approach was known mechanically by creating use of the fuzzy rule learning strategy, that are more practical for detecting intrusion in a very electronic network. At first, the exact rules were generated by mining the one portion of frequent things from attack knowledge to as traditional knowledge. Then, fuzzy rules was known by fuzzifying the definite rules and these rules get to fuzzy expert system, that classifies the knowledge. We have used KDD dataset for determining the performance of the proposed system and experiment results showed that the proposed technique is very effective in detecting various intrusions in computer networks.

The main attribute of fraud detection techniques is comparing network traffic with a defined or trained intrusion pattern in order to decide whether it is actually a attack or a genuine transaction. In case of fraud detection techniques involve any major deviation from a system of normal behavior. In Hybrid framework, we are using the advantages of both abuse and anomaly detection, and hence providing accuracy and speed to the detection mechanism. To improve the clustering or to modify the cluster into simpler we can modify the clustering algorithm by adding a weight of data features. The result proves that our framework produces a higher detection rate and low negative rate, compared to other framework. In the hybrid framework, to improve the performance of the fraud detection part, misuse detection is first applied to filter out the known intrusions from the given datasets. Thus, the number of connections in the fraud detection component is heavily reduced. The limitation of this hybrid system is that the dataset stored always gets very less for the new dataset. Another problem associated with the hybrid system, in fraud detection, some intrusions cannot be detected if it has a very high degree of closeness. In future, more advanced data mining algorithms could be investigated to bypass the previous limitations. And try to make all process online. The performance of weighted k means algorithm is strongly dependent on value of k clusters. Trying to find the best way to decide the value of k.

III. CATEGORIZATION OF CUSTOMER ACCORDING TO THEIR BEHAVIOUR

The customer's behaviors are categorized using the various parameters and then divided into 2 categories which are as follows:

A. Normal behavior

Means the customer is genuine and there is no suspicion detected by the expert system, so he can well proceed to the transaction with the normal flow.

B. Suspicious behavior

Means the customer is crossing the threshold of the parameters which is set for him and there is full chance of suspicious activities or fraudulent transactions. And the customer here have to go bypass the security check by the expert system.

IV. PARAMETERS USED FOR CHARACTERIZATION

For the first 10 transactions the bank will ask the OTP again and again to store some genuine values into the system. Based on these data only it will set some threshold or base value of the following parameters, failing which the customer havt to challenge the extra security of One Time Password for the successful transaction :

A. Amount

This is the most important parameters to detect whether the client is his customer or not, as if the amount of transaction is very high that it crosses the threshold generated by system, it may be a fraudulent activity.

B. Location

This is the second parameter to determine the originality of the user, suppose the coordinates of the user is very much deviated from his past 10 transactions, there is a sure chance

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

of someone hacking from sitting in far. The system thus generates the location and then categorizes it as normal or suspicious.

C. Mac Address

Suppose the user has till be making payments with 1 or 2 devices, the system will mark those devices as trusted one for him. And if the new transaction is being made by some random un-trusted device, fraudulent chances are very high.

D. No. of Successful Attempts

The number of successful attempt of login is also a key parameter as it shows whether the client is cracking the password by just hit or trail or is he a trustworthy customer.

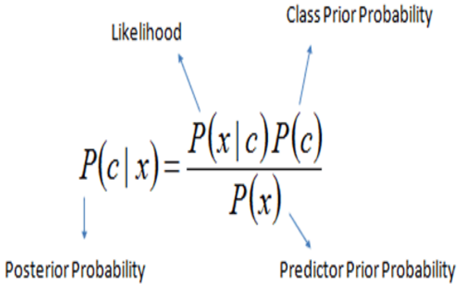
V. NAIVE BAYES BASED DEVELOPMENT

In this project, the Naive Bayes classifier is used to predict the outcome of events. Naive Bayesian algorithm is not just one algorithm but it is group of collection of classification algorithm which is based on Bayes theorem, this algorithm is stand alone algorithm but family of algorithm where all the algorithm share a common principle that every pair of feature which is classified is not dependent on each other.

In our project we have dataset of various parameter like location, transaction amount, type of transaction , trusted device, type of merchant, number of successful attempt to login into the system. Each of these classifies the condition as either safe or unsafe for doing the transactions. These dataset is divided into 2 parts which is feature matrix and response matrix . Feature matrix contain all vector which are dependent on each other while the other one contain the result set of the above tuples.

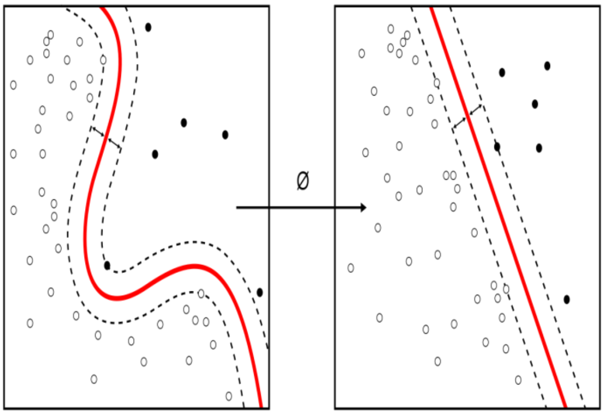
The fundamental property of naive algorithm is that each and every event is independent of each other for example in our case trusted device has no connection with having type of merchant , location and transaction amount, and each and every event has equal weight.

The value is calculated by mathematical function where probability of one event being occurred given that other event already occurred is given by the formulae below where A denotes the probabilty of event A to occur given that B had already occurred.

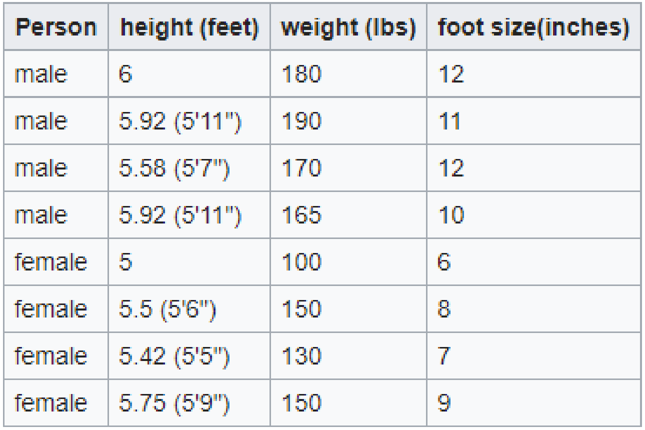


$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

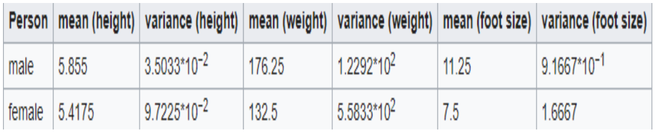
$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$



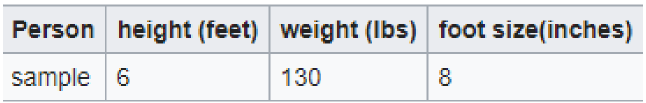
An Example of working of Naïve Bayes Algorithm.



Person	height (feet)	weight (lbs)	foot size(inches)
male	6	180	12
male	5.92 (5'11")	190	11
male	5.58 (5'7")	170	12
male	5.92 (5'11")	165	10
female	5	100	6
female	5.5 (5'6")	150	8
female	5.42 (5'5")	130	7
female	5.75 (5'9")	150	9



Person	mean (height)	variance (height)	mean (weight)	variance (weight)	mean (foot size)	variance (foot size)
male	5.855	$3.5033*10^{-2}$	176.25	$1.2292*10^{2}$	11.25	$9.1667*10^{-1}$
female	5.4175	$9.7225*10^{-2}$	132.5	$5.5833*10^{2}$	7.5	1.6667



Person	height (feet)	weight (lbs)	foot size(inches)
sample	6	130	8

$$\text{posterior (male)} = \frac{P(\text{male})\,p(\text{height} \mid \text{male})\,p(\text{weight} \mid \text{male})\,p(\text{foot size} \mid \text{male})}{\text{evidence}}$$

$$\text{posterior (female)} = \frac{P(\text{female})\,p(\text{height} \mid \text{female})\,p(\text{weight} \mid \text{female})\,p(\text{foot size} \mid \text{female})}{\text{evidence}}$$

VI. SOLUTION TABLE

Table 1.Trusted Device

	Yes	No	P(Yes)	P(No)
Trusted Device	2	1	2/5	1/3
Untrusted Device	3	2	3/5	2/3
Total	5	3	100%	100%

Table 2. Location

	Yes	No	P(Yes)	P(No)
Saved	3	2	3/7	2/8
New	4	6	4/7	6/8
Total	7	8	100%	100%

Table 3. No of Successful Attempt

	Yes	No	P(Yes)	P(No)
Success	1	2	1/5	2/4
Failure	4	2	4/5	2/4
Total	5	4	100%	100%

Table 4. Amount

	Yes	No	P(Yes)	P(No)
Silver	1	2	1/10	2/8
Gold	4	2	4/10	2/8
Platinurm	5	4	5/10	4/8
Total	10	8	100%	100%

VII. CONCLUSION

In this paper, using Naïve Bayes algorithm combined with different data mining algorithm generates a expert system, that will first get trained by the user for first 10 transaction and will create a dataset with some threshold values. Based on these data set and values the system will detect the behavior of the new transaction and categorize it as normal or suspicious, if normal no issues but if suspicious the user has to bypass the OTP challenge and other challenges if login attempt is failed too many times, this can include personal question appended by some unique pattern given by the bank at the time of making the account. This will create a belief in the customers that their money is safe even in clouds and they don’t have to worry doing any online banking transactions, this will create a good impression of bank also and the business of bank will reach sky high.

REFERENCES

1. S. P. Ketkar, R. Shankar , & D. K. Banwet, "Telecom KYC and mobile banking regulation", An exploratory study. Journal of Banking Regulation Advance online publication, 2013. http://dx.doi.org/10.1057/jbr.2013.
2. L. Wong, "Money-laundering in Southeast Asia", liberalism and govern mentality at work. Contemporary Politics, 19(2), 221-233, 2013. http://dx.doi.org/10.1080/13569775.2013.785832
3. M. Hepp, P. Leenheer, , A. de Moor, & Y. Sure, "Ontology management" semantic web, semantic web services, and business applications", Semantic Web and Beyond, Vol. 7. Springer, 2008. http://dx.doi.org/10.1007/978-0-387-69900-4
4. S. Raza, , & S. Haider, "Suspicious activity reporting using dynamic bayesian networks", Procedia Computer Science, 3, 987-991 , 2011. http://dx.doi.org/10.1016/j.procs.2010.12.162
5. J. T. S. Quah and M. Sriganesh, "Real-time credit card fraud detection using computational intelligence, Expert Systems with Applications, 3112, pp.17-9
6. L. Fang, M. Cai, H. Fu, and J. Dong, "Ontology-Based Fraud Detection," in Computational Science – ICCS 3112, pp.1048-1055, 2013.
7. D. Sanchez, M. A. Vila, L. Cerda, and J. M. Serrano, "Association rules applied to credit card fraud detection," Expert Systems with Applications, 2008, pp.1-14, 2008.
8. A. A Ramaki, R Asgari ,R.E Atani , "Credit card fraud detection based onontology graph ", International Journal of Security, Privacy and Trust Management (IJSPTM), Vol. 1, No 5, October 2012.
9. G. Montazer and L Saroukhani , "Design and implementation of a fuzzy expert system for suspicious behavior detection in e-banking system". 3.; 1 (1 and 2) :9-18 , 2009
10. Q Rajput, N.S Khan, A Larik , S Haider,"Ontology Based Expert-System for Suspicious Transactions Detection", Computer and Information Science; Vol. 7, No. 1; 2014ISSN 1913-8989 E-ISSN 1913-8997