Paper

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 3- Issue 2, Paper 10, August 2020

SECURED PAYMENT GATEWAY BY IDENTIFYING CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING ALGORITHMS

Ramya S S[1], Pon Sangeetha J[2], Yasin Javed S A[3], Mythili K[4]
[4] Assistant Professor
Department of Information Technology
Sri Krishna College of Technology
ramyasundaram72@gmail.com,17tuit103@skct.edu.in, 17tuit155@skct.edu.in, k.mythili@skct.edu.in

Abstract— *Internet has become one among the foremost important aspects in day-to-day life. Thanks to enormous amount of knowledge available on the web maintaining Confidentiality and Privacy of data transmitted across the network remains a problem. Many of us are unaware of the proper place to share their details .Lack of data results in hacking and misuse of private and tip by many intruders. to scale back this, a completely unique approach has been introduced, which analyses on all the sites within the network and approves only valid sites. Machine learning algorithms are utilized in the system to research the locations and help the users realize the integrity of the site. E-commerce being one among the main parts of the web, providing payment information within the payment gateway is also a challenging task. to supply authorization for the e-commerce transactions, this technique is developed which assures the integrity of site along side identification of fraudulence in payment gateway.*

I. INTRODUCTION

Ecommerce, also recognised as electronic commerce or internet commerce, refers to the shopping for and selling of goods or services the usage of the internet, and the transfer of cash and statistics to execute these transactions. One can get a lot of benefits by way of opting for eCommerce as it offers a complete vary of benefits to outlets and merchants. Electronic Commerce is also recognized as e-commerce that consists of the buying and promoting of merchandise or offerings through electronic structures like pc networks and the Internet.

Modern electronic commerce normally makes use of the World Wide Web for at least one section of the transaction's life cycle even though it may also use different technologies such as e-mail. Typical e-commerce transactions consist of the purchase of on line books and tune purchases , and to a terrific deal a good deal much less extent, customized/personalized on line liquor hold stock services. There are three areas of e-commerce: online retailing, digital markets, and online auctions. E-commerce is supported by means of electronic business.

Credit card fraud is a term for theft and fraud committed the usage of a payment card, such as a credit card or debit card, as an illegal source of money in a transaction. So, can also be to attain products by fraudulent paying, or to attain unauthorized money from an account.

The credit card fraud are restrained to about 0.1% of all card transactions, but it meets large economic losses as the illegal transactions have been enormous money transaction. It is considered to be 0.04% of all monthly active accounts are fraudulent. Even many fraud detection algorithms are designed to prevent fraudulent transactions, many transactions processed which leads to millions of dollars in losses.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 3- Issue 2, Paper 10, August 2020

The theme of credit card fraud is for illegal transactions and if a card is lost or stolen, it is possible to block on that account by issuing bank. Mostly, all bank services supports to encourage instantaneous reporting on fraudulent. In case the bank takes time to cancel the card, the thief can make illegal transactions on that time. If the thief is a hacker, they particularly try to hack the payment gateway on that account and make illegal transactions or purchases.

The adequate safety measure on most of the cards is a signature panel, but it is not much secure since the actual design of a signature can also be forged. Some merchants need to see photo ID on the credit card details itself, for the verification of the credit card, so can be able to avoid fraudulent. Before giving payment information on the website, it is needed to check whether the payment website is legitimate or not. Because there is a lot of illegal website available which can easily hack our account. With database hacking they can use our account illegally, so may easily buying products without our knowledge.

Skimming, which is also a type of credit card fraud where the thief use a small device which is placed over any Card Reader. So the skimmer can get credit card details and payment information when cardholder swipes a card. In Pharming technique, the fraudsters develop or use the fake websites which is appears to be original website. By this technique, credit card, debit card and payment information is stolen.

II. LITREATURE SURVEY

ID3 J48 is a supervised machine learning algorithm which uses decision tree concepts. WEKA tool provides java application environment which enhances this C4.5 algorithm. This algorithm finds the observed samples and final values through internal nodes and final nodes of the decision tree. It has many features for finding accuracy such as pruning, derivation rules, attribute ranges etc. From this algorithm we can check the accuracy of the data without interpretation.

In ensemble, various models are trained a single problem for getting better accuracy than single trained model. It has some techniques which can be applied for more accuracy. In bagging, collects more subsets of trained data from given sample. This collection of n subsets are trained on their decision tree algorithm. Hence predictions from various decision trees are used to find more accuracy. For supervised learning, decision tree creates a graphical representation of a tree which enhances both classification and regression. It has pruning feature which eliminates the unwanted nodes which is not much important. [1]

The deep learning-based model performs better than other algorithms since it has efficient features for predicting large amount of data. It can analyse the data through data analysis and works on both machine learning algorithms and artificial intelligence. Most of these algorithm take non linear inputs and transforms the input of hierarchy into statistical output.[2]

Credit card fraud is categorized into two types, they are behavioural fraud and application fraud. Application fraud is rare, it occurs when spending a lot of money at short period of time with fake personal information. Most of the time, credit card fraud is behavioural fraud which occurs when details of the cards are stolen and make fraudulent transactions. In this paper, the algorithms are enhanced for Behavioural fraud Transactions and Deriving Features, Updating accuracy score of set of classifiers of cardholder[3].

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 3- Issue 2, Paper 10, August 2020

Various examinations have recommended that the group methodology could be a promising way to deal with improve arrangement execution and to acquire an increasingly steady subset of highlights. Likewise investigating the advantage of dataset dividing idea and nature of the classifier Consequently, these investigations makes us to investigate the Half and half Group Highlight Choice (HEFS) structure, which comprises of two cycles, to be specific, information irritation cycle and capacity annoyance cycle. Randomisation alludes to the rearranging of the example lines, and is used to guarantee that the examples in each parcel are progressively adjusted.

A component that is genuinely prescient is most ensured to exist in all dataset segments, in this manner empowering it to be chosen (through the crossing point activity) as a component of the optional element subset. This finishes up a solitary information irritation cycle. The function bother troupe part totals its sources of info through an association operation. Through the capacity irritation cycle, the knowledge of various channel measures can be utilized, along these lines prompting the pattern list of capabilities that is less vulnerable to over fitting. It is profoundly alluring to devise another methodology that is more adaptable to decide the ideal cut-off position.[4]

Three techniques for predicting credit card fraud:

A. Support Vector Machines

SVM algorithm separates the data into two categories positive instance and negative instance through hyperplane using support vectors. .Here, particularly the legal and illegal category is predicted.

B. Naïve Bayes

Naïve Bayes algorithm is a classification technique, has a feature of multi class prediction by assumes random feature in one class is independently on any other feature of the class.

III. PROPOSED SYSTEM

We implement the secured payment gateway by using three machine learning algorithms

1) **Decision tree**-most simple, comprehensive, specific easy to use, versatile and a supervised ML algorithm used to solve classification and regression problems.
2) **KNN**- is both supervised machine learning algorithm based on predictions.
3) **AES**- a machine learning algorithm used for encryption and decryption**.**

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 3- Issue 2, Paper 10, August 2020

MODULE 1: DECISION TREE

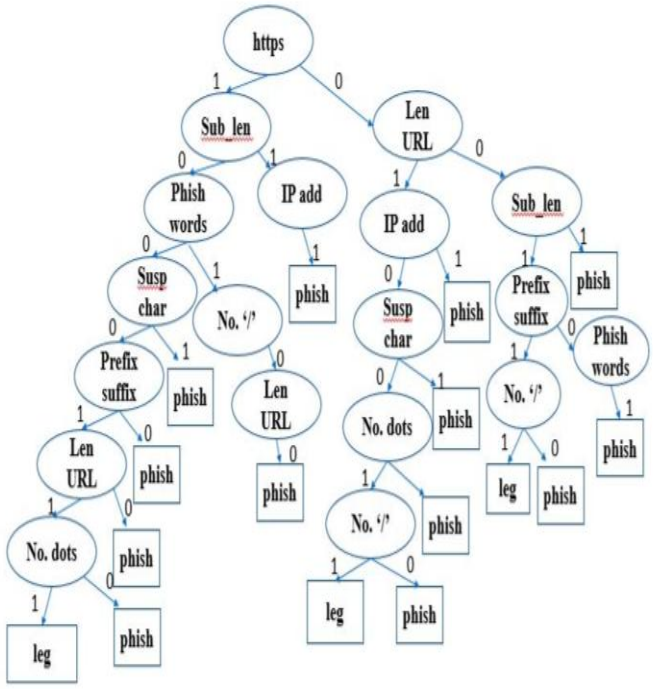

Fig. 1 Decision Tree

1. Preprocessing State

Step 1: The system is given prior information about the URL of the phishing and legitimate website.
Step 2: The feature extractor extract feature values from URL based on prior information. The a variety of features extracted are unique address(IP), Length of URL, Prefix and Suffix, variety and size of subdomain, phishing words, range of '/' and HTTPS protocol. The values of these features are given as follows based on one of a kind conditions:

- Unique address: If the URL consist of an unique address then the value is assigned as 1 or vice versa.
- Length of URL: When the URL length exceeds the 54, the value assigned is 0 else the value is 1.
- Mysterious Characters: If a mysterious character is present in the URL then the value is 1 else the value is 0.
- Prefix and suffix: If prefix and suffix are present in the URL then the value is 1 else the value is 0.
- Number of subdomain: If the number of dots is less than three then the value is assigned as 1 otherwise the value assigned is 0.
- Length of subdomain: If the length of subdomain in URL is greater than five then value assigned is 0, otherwise value assigned is 1.
- Phishing words in URL: If the URL consists of phishing terms then the value is 1 otherwise the value is 0.
- Number of '/': If the number of '/' is less than 5 then the value is assigned as 1 If it exceeds 5 then

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 3- Issue 2, Paper 10, August 2020

value is 0.

- HTTPS protocol: If the URL has https protocol then value assigned is 1 otherwise 0.

Step 3: The extracted features and their corresponding values are stored in a file and used as an input and passed to the classifier generator, which generates set of rules by using the input features and the C4.5 algorithm.

2. Detection Phase

Step 1: If a user request for a page the requested site URL is received by the system.
Step 2: The URL is then given to the feature extractor, which extracts the feature values through the previously defined URLbased features.
Step 3: The feature values act as an input to the classifier.
Step 4: The classifier evaluates whether the requested site is trusted or not based on prior information. If the site is classified as untrusted then the site is labeled as phishing by the classifier then the user is made alert about the classification result.

MODULE 2: KNN

A. Behavioural fraud

While shopping any item from online we submit the card information of any credit card without the knowledge of owner thereby behavioural fraud takes place .Fraud bar ways that two like Mastercard validation etc. However, fraudsters square measure adaptive and square measure obtainable up with several misleads and concepts to interrupt the prevailing bar mechanisms. Once the fraud bar techniques square measure unsuccessful, there is the necessity of developing associate economical fraud identification system (FIS) for identifying master card thefts and maintaining the feasibility of the payment system.[6]

B. KNN

KNN is a considered as both non-parametric and an illustration of lazy learning.

Non-parametric implies that it makes no presumptions. The model is made up altogether from the information given to it instead of expecting its design is ordinary.

Lazy learning implies that the calculation makes no speculations. This implies that there is small preparing included when utilizing this strategy. Along these lines, the entirety of the preparation information is additionally utilized in testing when utilizing KNN.

KNN for Classification

At the point when KNN is utilized for characterization, the yield can be determined as the class with the most elevated recurrence from the K-most comparable occurrences. Each occurrence fundamentally votes in favor of their group and the class with the most votes is taken as the expectation.

Class probabilities can be determined as the standardized recurrence of tests that have a place with each class in the arrangement of K most comparative examples for another information occasion. For instance, in a paired order issue (class is 0 or 1):

p(class=0)=count(class=0)/ (count(class=0)+count(class=1))

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 3- Issue 2, Paper 10, August 2020

In the event that you are utilizing K and you have a significantly number of classes (for example 2) it is a smart thought to pick a K incentive with an odd number to evade a tie. Also, the converse, utilize a much number for K when you have an odd number of classes.

Ties can be broken consistently by expanding K by 1 and looking at the class of the next most similar instance in the training dataset.

C. Proposed Methodology

Step 1: Select the number K of the neighbors

Step 2: Calculate the Euclidean distance of K number of neighbors

Step 3: Take the K closest neighbors according to the determined Euclidean distance.

Step 4: Among these k neighbors, check the quantity of the information focuses in every class.

Step 5: Assign the new information focuses to that class for which the quantity of the neighbor is most extreme.

Step 6: Our model is prepared.

MODULE 3: ENCRYPTION

Transmitting steer like plain text countersign via a wire is continually at risk of security. We can encrypt and trust SSL to transfer the secured data. In this, AES (Advanced Encryption Standard) symmetric encryption algorithm in java with CBC mode which is much more efficient than DES algorithm is used:

AES=3(DES)

It is further classified into Asymmetric and Symmetric encryption. Asymmetric encryption such as RSA includes two distinctive keys as public and private keys. We can safeguard our data by encrypting it with public key and a matching private key is used to decrypt the same. Asymmetric encryption is preferred when there are 2 different terminals are concerned. Another encryption method, Symmetric encryption recognised a private key or secret key to encrypt and decrypt sensitive information. This encryption type is much more efficient than asymmetric encryption, and can used in systems such as database system. Some examples of symmetric encryptions are Two fish, Blowfish, 3 DES, AES.

For encryption, we can enter the data format that we desire to encrypt. The AES engine rely on a plain-text and a secret key for encryption and same secret key is required again to decrypt it. Now pick the block cipher mode of encryption. ECB is the easiest encryption mode and does not require IV for encryption. The input simple text will be segmented into blocks and each block will be encrypted with the key value supplied and for this reason identical simple text blocks are encrypted into identical cipher text blocks. CBC mode is essential and it requires IV to make every information unique. If no IV is entered then default will be used right here for CBC mode and that defaults to a zero based byte to avoid errors.[9]

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 3- Issue 2, Paper 10, August 2020

IV. ARCHITECTURE

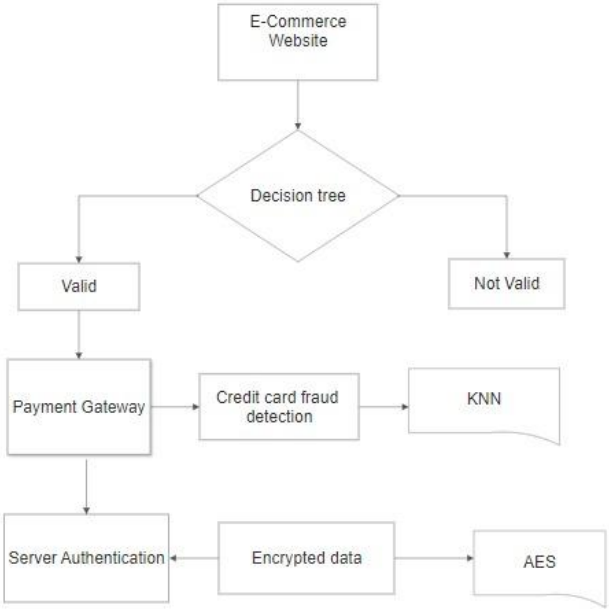

Fig. 2 Architecture

V. CONCLUSION

The proposed system uses multiple algorithms to validate the user in financial transactions. The system easily analyses the various parameters for validating an authenticated user in a network and their credentials. So, even at early stage it is easy to detect the fraudulence in the network. Loss of data, access of confidential data can be avoided using this algorithm. The systems shows a performance increase compared to many other techniques used in the network.

REFERENCES

[1] A Smart Methodology for Analyzing Secure EBanking and E-Commerce Websites Rana M. Amir Latif*, Muhammad Umer, Tayyaba Tariq, Muhammad Farhan, Osama Rizwan, Ghazanfar Ali Department of Computer Science COMSATS University Islamabad, Sahiwal Campus Sahiwal, Pakistan {ranaamir10611*, muhammadumer063, tayyaba.tariq.tt, farhansajid, rizwan.osama.official,ghazanfarali78622}@gmail.com.

[2] Champion-challenger analysis for credit card fraud detection: hybrid ensemble and deeplearning Eunji Kim, Jehyuk Lee, Hunsik Shin, Hoseong Yang, Sungzoon Cho, Seung-kwanNam,Youngmi Song, Jeong-a Yoon, Jong-il Kim PII: S0957-4174(19)30216-7

[3] Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism Changjun Jiang, Jiahui Song, Guanjun Liu , Member, IEEE, LutaoZheng, and Wenjing Luan, Student Member, IEEE

[4] A New Hybrid Ensemble Feature Selection Frameworkfor Machine Learning-based Phishing Detection System Kang LengChiewa, Choon Lin Tana,_, KokSheikWongb, Kelvin S.C. Yongc, Wei King Tionga a Faculty of Computer Science and Information Technology, UniversitiMalaysiaSarawak, 94300 Kota Samarahan, Sarawak, Malaysia b School of Information Technology, Monash University Malaysia, 47500 Bandar Sunway, Selangor, Malaysia c Department of Electrical and Computer Engineering, Faculty of Engineering and Science, Curtin University, CDT

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 3- Issue 2, Paper 10, August 2020

250, 98009 Miri, Sarawak, Malaysia.

[5] Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier MasoumehZareapoora , PouryaShamsolmoalia,b -a Department of Computer science, Jamia Hamdard University, New Delhi, India b Department of Computer Science, Baghin University, Kerman, Iran

[6] https://sift.com/sift-edu/prevent-fraud/avs-cvv2

[7] https://towardsdatascience.com/everything-you-need-to-know-about-neural-networks-and-backpropagation-machine-learning-made-easy-e5285bc2be3a?gi=a4423cd50665

[8] DeepSentiment: Finding Malicious Sentiment in Online Social Network based on Dynamic Deep Learning Putra Wanda1,2, Huang Jin Jie1 , Member, IAENG

[9] https://en.wikipedia.org/wiki/Block_cipher_mode_of_operation

[10] https://www.ssl2buy.com/wiki/symmetric-vs-asymmetric-encryption-what-are-differences