

Using Associative Classification for Detecting E-Banking Phishing

Nwachukwu C.B.

Department of Computer Science, Ignatius Ajuru
University of Education, Rivers State, Nigeria

N.A. Ojekudo

Head of Department, Department of Computer
Science, Ignatius Ajuru University of Education,
Rivers State, Nigeria

ABSTRACT

Phishing attack has become very common in recent years especially in the financial application setting. A lot of losses have been recorded as a result of this attack from users and subscribers all over the globe. The research community has also been very concerned about this development with leaves an unsavoury aftermath on its victims, hence, several models, systems, architecture and frameworks have been developed by the researcher in an attempt to tackle this menace. The key limitations of these developments include lack of standard classification, low model efficiency and performance in terms of speed and time and high cost of developing the models. In this work, we have developed an enhanced Phishing detection system using Associative Classification technique. The Structured System Analysis and Design Methodology (SSADM) was adopted in this approach. The system was implemented using Hypertext Preprocessor (PHP) and MySQL as database. From our results the proposed model had an overall accuracy score of 86.6% which outperformed the existing system with 55.9% when evaluated using selected parameters. This system could be beneficial to Nigerian Banks, to Digital Banking Application Users and to the entire research community.

Keywords

Phishing, Associative Classification, Malware, Identity Theft

1. INTRODUCTION

Digital communications is the new norm in our world today. Many activities which were formerly manual and required user presence have been integrated into the digital platform. Such activities include banking, shopping and other forms of exchange of goods and services.

The introduction of the digital platform met a welcoming acceptance by large volume of people all over the world. It also brought about ease of doing things from the convenience of one's comfort zones. However, this invention has also met with

great cyber opposition by attackers who pose a serious threat to the subscribers to these platforms.

Phishing is an example of a cyber-attack in which an attacker attempts to obtain sensitive information or data such as usernames, passwords, credit card numbers, bank account numbers or other vital financial details of a person by posing to be a trustworthy entity or a legitimate source. In Nigeria, this is commonly perpetrated via voice calls, whereby an attacker randomly calls a subscribers, provides details such as the name(s), name of banks and other details of the person (just to sound more convincingly true) and request that the person should provide his/her account details or other financial details which can be used to access funds in the person's account under the guise of either an upgrade or some sort of verification of the person's account. Before the information about these attackers spread widely, so many people were already victims of the attack. Huge sums were reportedly lost etc.

Phishing could also come via emails inform of an unprompted request to verify certain details of a person's financial status posing to be a trusted services. Some attackers can clone a particular website and change little data in the URL and use it to send malicious links and mails to persons randomly which upon clicking saps personal and vital information about the person. SMSs, Direct messages on social media accounts are also some channels for phishing attack. There are also other forms of phishing which include email phishing, website phishing, spear phishing, Whaling, tab napping, evil twin phishing etc.

Associative classification is a machine learning approach that aims to build accurate, efficient and compact classification models/classification by combining paradigms for classification and associate rule discovery systems.

It is a scientific study that is being used by decision systems. Some common algorithms used for the associative classification include: Naïve Bayes, K-NN, VM etc.

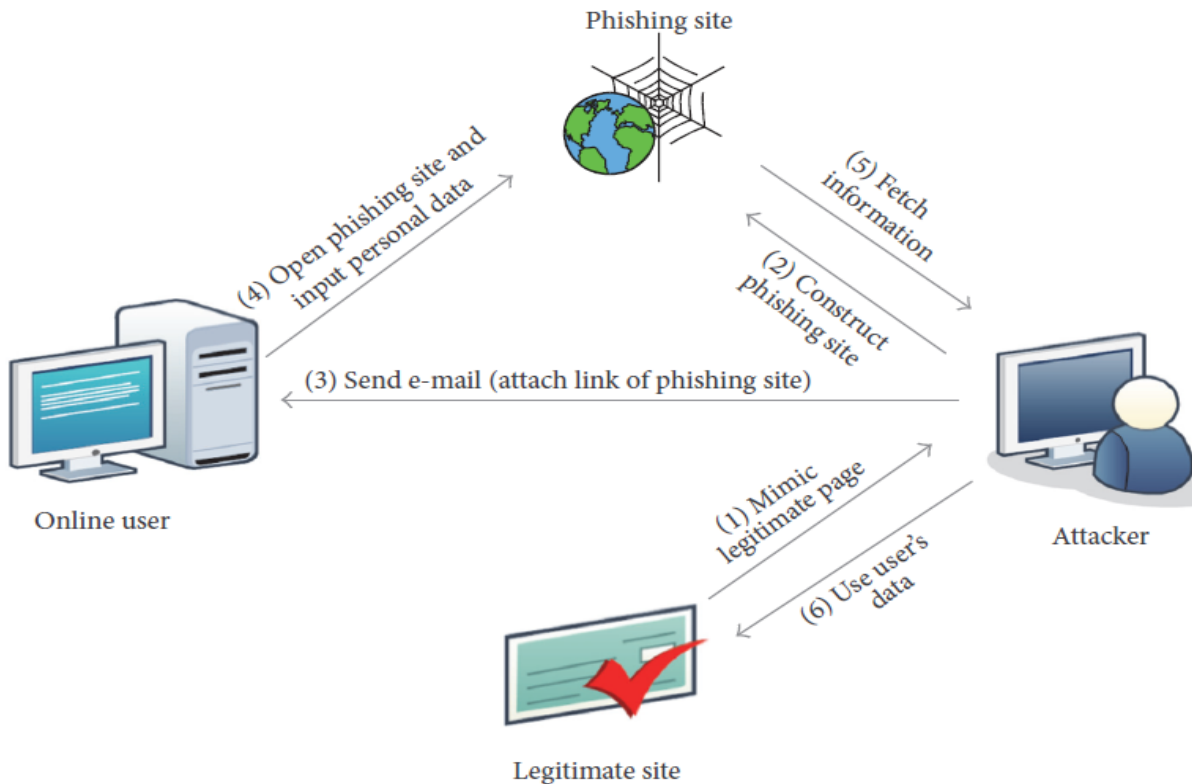


Fig. 1: Demonstration of the Phishing Mechanism [2]

Due to the level of impact caused by phishing activities researchers have developed several models for detecting, predicting or blocking the attack. One of these models is the use of associative classification.

However, in this study we will adopt Support vector machines (SVMs) for our classification and detection process.

1.1 Statement of Problem

Phishing detection models currently in place have certain limitations which create a vacuum for attackers to still penetrate even with these models in use. Some of the limitations include lack of standard classification, low model efficiency and performance in terms of speed and time and high cost of developing the models.

1.2 Aim and Objectives

The aim of this study is to develop an e-Banking Phishing model using associative classification. The specific objectives are to:

- i Simulate a secure e-Banking platform for detection of phish using associative classification.
- ii Implement using Hypertext Preprocessor and MySQL as database.
- iii Compare our results with other existing system performances.

2. RELATED WORKS

In this section, we are going to review some of the efforts made by researchers in trying to tackle these attacks.

Ajlouni et al [1] presented detecting phishing websites using associative classification. They investigated the use of automated data mining techniques for the detection of complex

phishing problems. They adopted two associative classification algorithms called CBA and MCAR which they applied on phishing datasets. Their results showed that MCAR outperformed other algorithms. However, their model could not be used to block the malicious attack, only for detection.

Ratnaparkhi and Jambhulkar [2] proposed a framework for detection and prevention of phishing website using machine learning approach. They developed a novel feature approach based on Recurrent Neural Network (RNN) which was scalable and proactive. Classification approaches were employed to classify legitimate and illegitimate messages which were trained by the RNN. However, their framework could not automatically detect fake sites or URLs.

Pawanet al.[3] presented robust and efficient phishing detection via page component similarity. They outlined two ways to detect phishing websites by the blacklist. The first way includes five heuristics to enumerate simple combinations of known phishing sites to discover new phishing URLs. The second way consists of an approximate matching algorithm that dissects a URL into multiple components that are matched individually against entries in the blacklist.

Futai et al [4] proposed graph mining technique for web phishing detection. Their model recognized some possible phishing which URL assessment could not distinguish. They used gathering association among customer and site to get dataset from the authentic traffic of a Large ISP. They combined the gathering association outline with AD and URL, called AD-URL Graph and the Phishing site was detected through the common direction of the chart.

Armano et al [5] proposed utilization of extra feature in the program which is Real-Time Client-Side Phishing Prevention. They made use of data collected from sites the customer had

visited in the event that it is a phish and caution the client. Additionally, they outlined the goal of the phish and the offers to divert the client there. An admonition message was displayed in interface while the foundation showed the phishing page obscured by a dark semi-straightforward layer preventing alliance with the site.

Jain and Gupta [6] proposed analysis of visual similarity based approaches for phishing detection. A detailed description of phishing attacks and some of the most recent visual similarity approaches for phishing detection and its comparative study was provided.

Aburrous et al [7] presented modeling intelligent phishing detection system for e-banking using fuzzy data mining. The central focus of their study was to beat the fuzziness of the banking system and in turn identify e-banking phishing websites. The model combined fuzzy logic and data mining algorithms to classify website factors and phishing types and characterize the attack criteria for the site. The proposed e-banking phishing website model showed the significance importance of the phishing website two criteria's (URL & Domain Identity) and (Security & Encryption) in the final phishing detection rate result.

Kazi et al [8] proposed phishing websites detection using associative banking. They proposed techniques that could detect and block phishing attacks via emails. The system was an end-user application which employed URL and hyperlink features with digital signatures to prevent the attack. The application acted as an interface between e-mail service and a user to provide secure communication between them.

Ramya et al [9] proposed an intelligent resilient and effective model that employed a new class based associative classification algorithm which is an advanced and efficient approach than all other association and classification Data Mining algorithms. This algorithm was used to characterize and identify factors and rules in order to classify the phishing website and their relationships. The searching space was reduced using class label as a rule mining step. The algorithm also harmonized the rule generation and classifier building phases, reducing the rule mining space when developing the classifier to help speed up the rule generation.

Alhamad et al [10] proposed detecting e-banking phishing website using C4.5 algorithm. They used the C5.4 algorithm for classification in the WEKA program, by analyzing data and classifying fake and legitimate sites to reduce the problem of phishing in banking services. After processing the data and applying the algorithm, the accuracy rate reached by the algorithm was 98.11%, while the algorithm's error rate was only 1.89%. However they could not deploy it on the internet and search engines to block phishes.

Mohammad et al [11] proposed an intelligent rule based phishing websites classification. They distinguished between a legitimate and a phishing websites using their features as parameters. They used 17 parameters in total. They also developed a feature for each parameter which was used for detecting its legitimacy. They employed several algorithms using WEKA such as C4.5, RIPPER, PRISM and CBA. Their results showed that C4.5 outperformed the other algorithms. It also showed that the most prominent feature for phishing detection was the Request URL.

Kamble et al [12] proposed detecting e-banking phishing

websites using Naïve Bayes classifier. They combine association and classification data mining algorithms to develop a model for detecting phishing in websites. The results demonstrated the possibility of using associative classification in real application to detect phishing. However, they did deploy their model for real world application and problem solving.

Divya and Vivithat [13] proposed phishing websites detection using associative classifiers. The main goal of their study was to experiment the use of automated data mining techniques in detecting complexities associated phishing. They adopted MCAR and CBA as their associative learning algorithms. Their results demonstrated that these algorithms outperformed the traditional techniques formerly in place.

Wankhede et al [14] proposed detecting the phishing websites using enhance secure algorithm. Their research focused on using a URL-based indicator for phishing attack. The finding of their research demonstrated that the most efficient way of prevention was to adopt awareness approach and educate users on how to spot fake URLs and avoid them. However, they could not include automatic detection of the page and compatibility with web browsers.

Damodaram and Valarmathi [15] proposed phishing websites detection using particle swarm optimization. They adopted this approach to overcome the limitation in previous studies such as time complexity and uncertain time convergence during classification. However, this study did not use association rule to classify phishes and thereby proffer a more efficient result.

3. ARCHITECTURE OF THE EXISTING SYSTEM

The existing system was proposed by Ratnaparkhi, and Jambhulkar [2]. The system was built in three modules consisting of the user i.e. he account holder, the bank system and the database. The database was used to collect details such as customer's personal details etc. the admin is generally in control of user authentication and ensures unauthorized access to the platform.

The existing system was responsible for giving access to user by simply matching anomalies, if matched then "Ok" otherwise check for identity. "Feature-Based Phishing Detection Technique", was used as a method in this system based on utilizing RNN machine to perform the classification. This method extracted and formed the feature set for a webpage. It uses a RNN machine as a classifier which has two phases, training phase and testing phase, during training phase it extracts feature set and while testing it predicted the website is legitimate or a phishing.

Experiments were carried out using a system with this configuration: Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz, 4GB memory, Windows 7, MySQL 5.1 backend database and JDK 1.8. The application was web-based used tool for design code in Eclipse and execute on Tomcat server.

Each module in their architecture worked independently. Each module watched the activities of the browser. It used a dual approach. IP address check as well as web page content analysis.

Each time the user enters a URL in the address bar, the URL was checked with a database of phishing sites. If a matching entry was not found in the database, then the source code of the page was checked for non-matching URLs, IP-based URLs, etc. When a URL is entered in the address bar, the browser monitor

is invoked. The browser monitor then queries the phishing database to check whether or not a match is found. If a match was not found then the browser monitor invokes the feature detection engine.

However, the method of detection of the URLs is not automatic but is manually controlled by the Admin therefore the system could still be penetrated by an attacker if the Admin is not on sit.

4. ARCHITECTURE OF THE PROPOSED SYSTEM

The proposed system is an enhancement of the existing system.

The proposed system contains two modules (user and banking module). The proposed system also contains a classification block which classifies URL features as either fake or real. The URL dataset was developed locally for the purpose of this research. The dataset contains about 30 URLs which were gotten from existing well known sites and also some fake URLs sent to our mails in the past were also scrapped and attached.

The SVM converts non-separable problems into separable problems by adding more dimensions to it. It makes SVM more powerful, flexible and accurate [2]. This classification algorithm was used to classify the URLs in the dataset as either fake or real by adding extra feature such as the authenticity tag to the URL after classification.

The Artificial Neural network (ANN) was used to train the data by assigning weights to the data for automatic detection in the future.

The database was used to hold user's details, phish classification results, updates and Phish URLs.

The advantages of the proposed system include:

- i Efficient classification model with minimal errors.
- ii Elimination of time and cost complexities.
- iii Good data training algorithm for better detection results.
- iv Good performance evaluation results from implementation.
- v User friendly design for easy navigation.

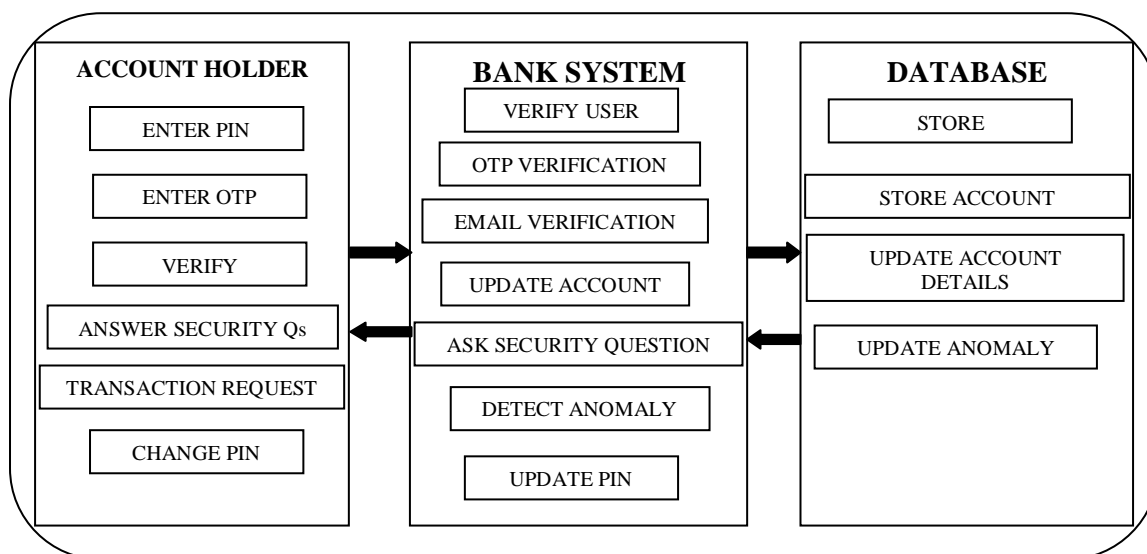


Fig. 2: System Architecture of the Existing System [2]

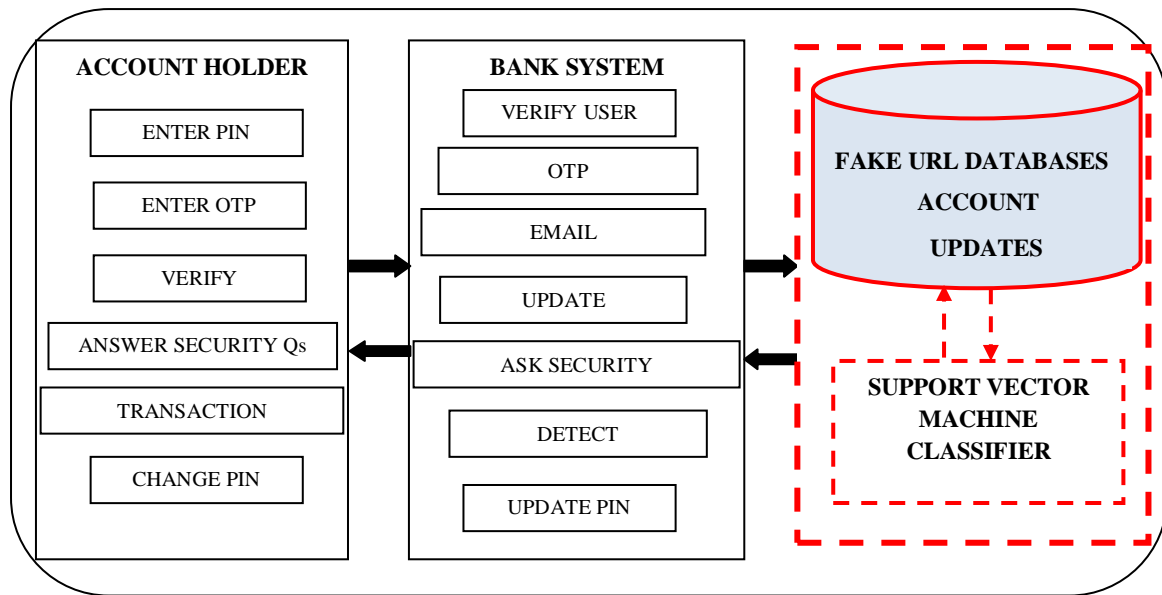


Fig. 3: Architecture of the Proposed System

4.1 ALGORITHM OF THE PROPOSED SYSTEM

The following steps will be followed to achieve the verification and blockage process.

- STEP 1: CREATE DATASETS OF URLs.
- STEP 2: INITIATE TRAINING USING ANN
- STEP 3: LAUNCH SVM CLASSIFIER
#DEFINE URL FEATURES
- STEP 4: IF URL != FEATURES
CLASSIFICATION = PHISH
- ELSE
CLASSIFICATION = LEGITIMATE
- STEP 5: INPUT URL AS VERIFICATION INPUT
- STEP 6: SCAN PHISH-BANK FOR URL.
- STEP 7: IF MATCH?
FLAG PHISHING ATTEMPT.
BLOCK URL.
- ELSE
SAFE SITE. CONTINUE
- STEP 8: UPDATE PHISHING BANK.
- STEP 9: END PROCESS

5. IMPLEMENTATION OF THE PROPOSED SYSTEM

The system was implemented using hypertext preprocessor (PHP) programming language and MySQL as database. The system was run locally using a local server (XAMPP). HTML was used to build the framework for the system. The system developed in this research is a simulation of the actual system. It is an experiment of how a real phishing detection system actually works.

The system specification includes:

- RAM:** At least 1 GB
- IDE:** Notepad++, Sublime Text
- HDD:** 500MB
- PROCESSOR:** at least 600-MHZ
Pentium III-class

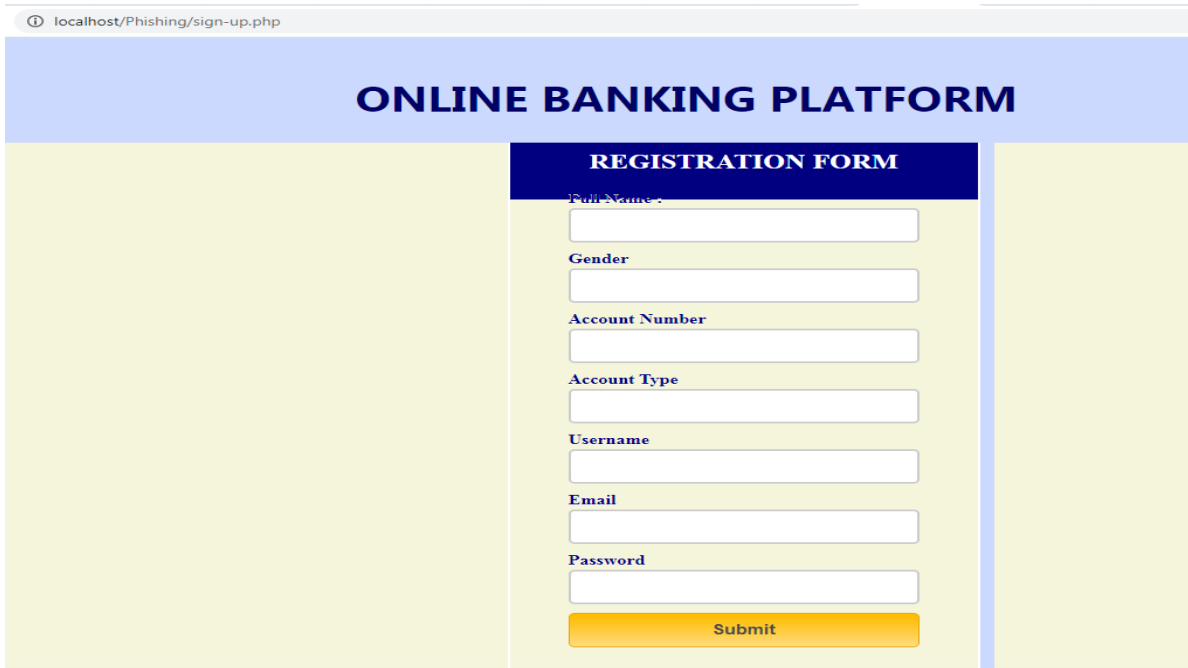
5.1 METHODOLOGY

The methodology adopted for the proposed system is the Structured System Analysis and Design Methodology (SSADM). This methodology was adopted because it helps in the development of expert systems in the process of replacing the existing system with a new one.

6. DISCUSSION OF RESULTS

The outputs and results from the performance evaluation of the proposed system are displayed in this section. For the basis of encryption, MD5 encryption algorithm was used to secure the user details stored in the database to ensure they are not retrieved for malicious purposes.

The proposed system has a registration and login page which enables only authorized users to gain access into the system.



The screenshot shows a web browser window with the address bar displaying 'localhost/Phishing/sign-up.php'. The main content area has a light blue header with the text 'ONLINE BANKING PLATFORM' in bold, dark blue letters. Below the header is a registration form with a dark blue title bar that says 'REGISTRATION FORM'. The form contains several input fields: 'Full Name', 'Gender', 'Account Number', 'Account Type', 'Username', 'Email', and 'Password'. Each field is a white rectangle with a thin border. At the bottom of the form is a yellow button with the text 'Submit' in black.

Fig. 4: Homepage of Digital Banking Platform

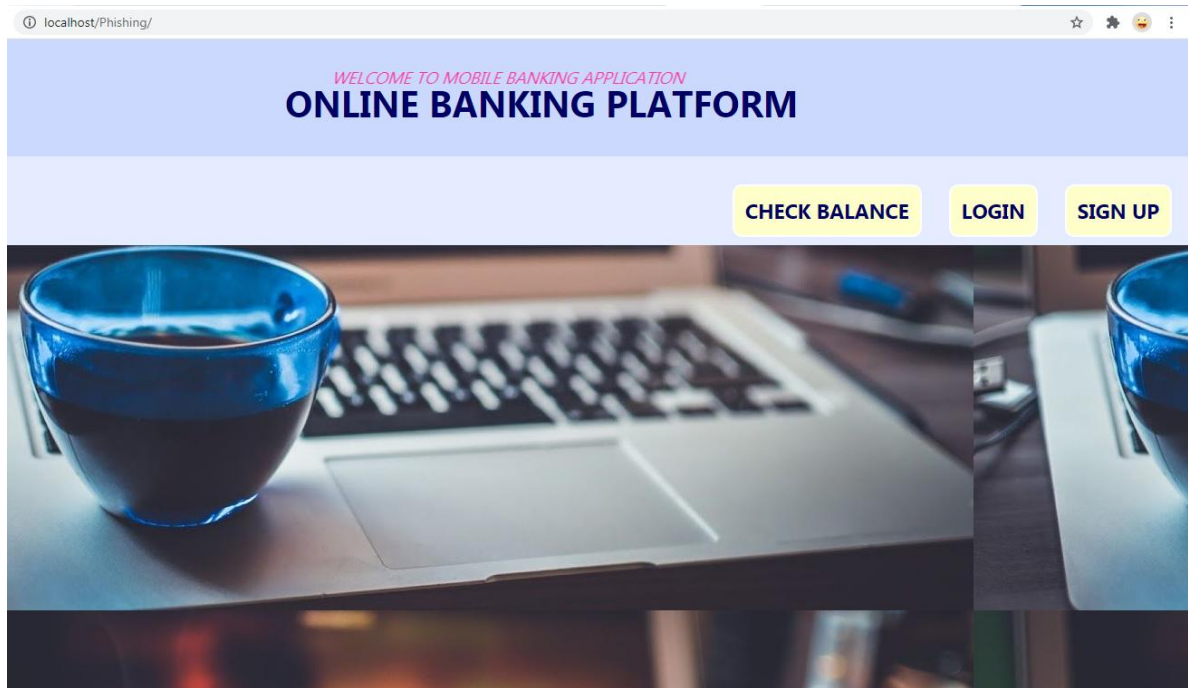


Fig. 5: Registration Page for New Subscribers

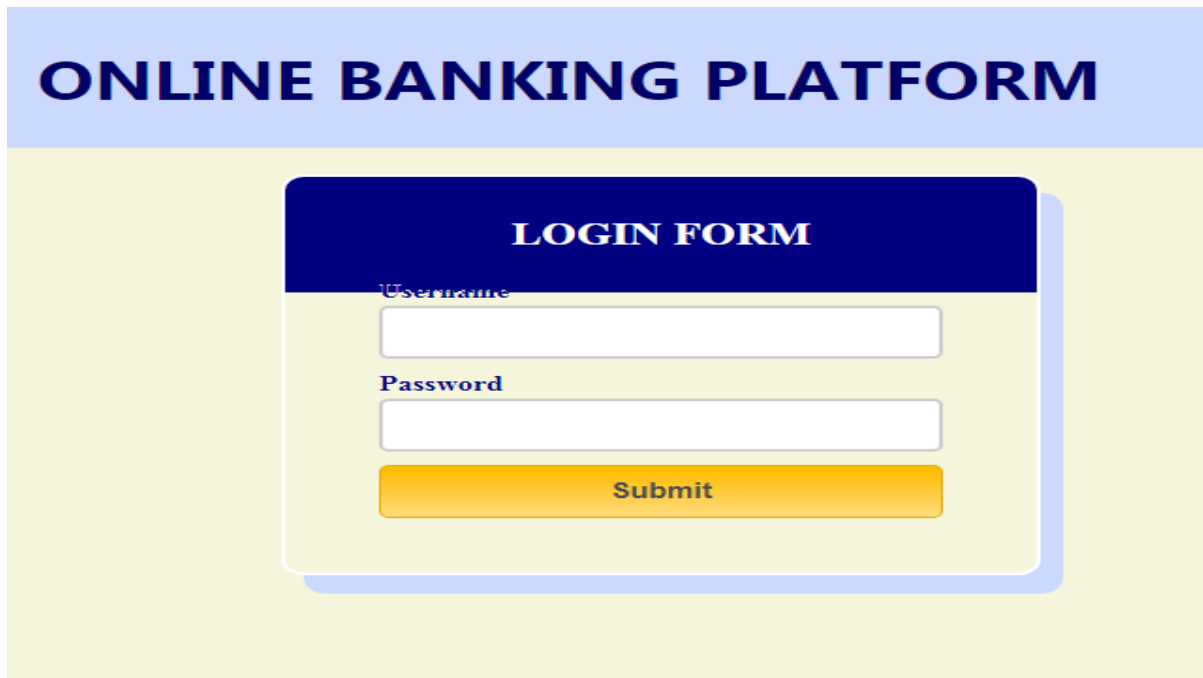


Fig. 6: Login Page to Banking Platform

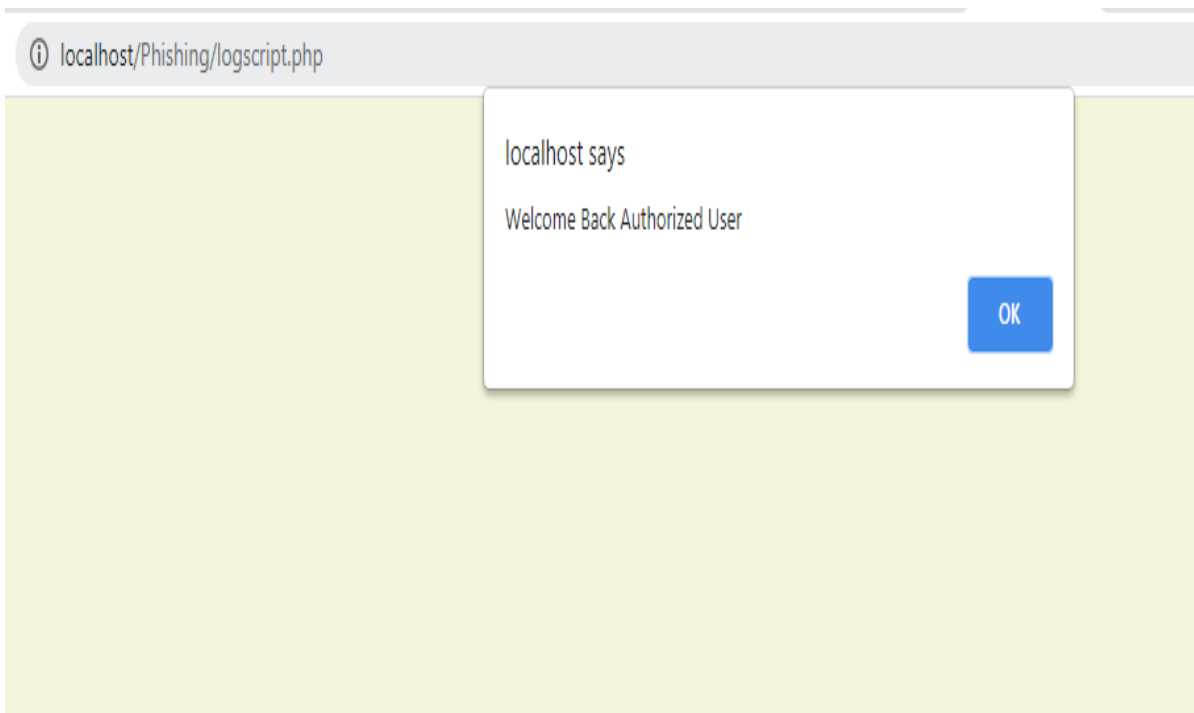


Fig. 7: User Authentication Alert

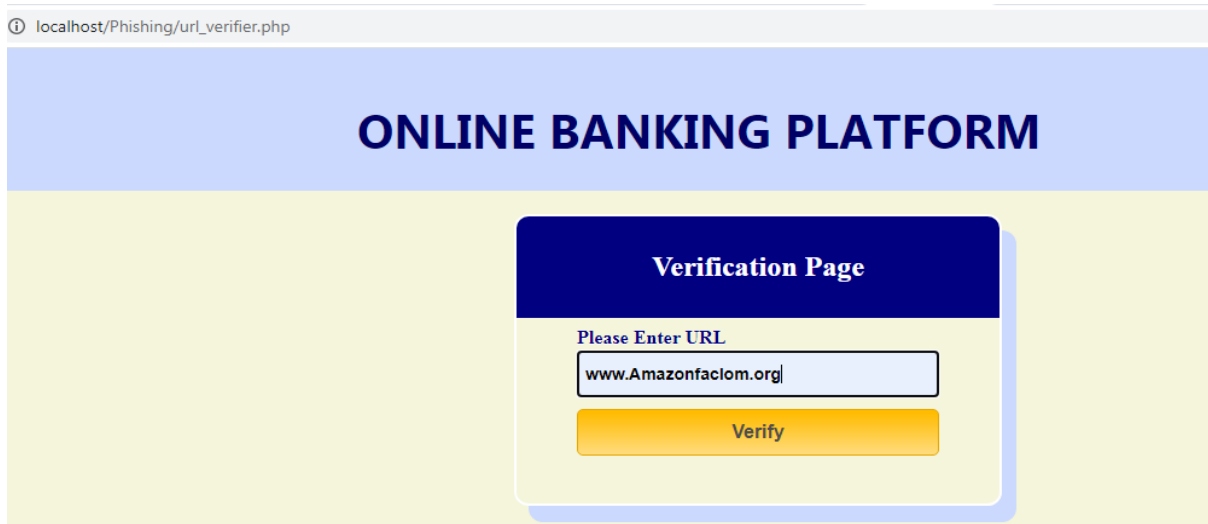


Fig. 8: URL Verification Prompt

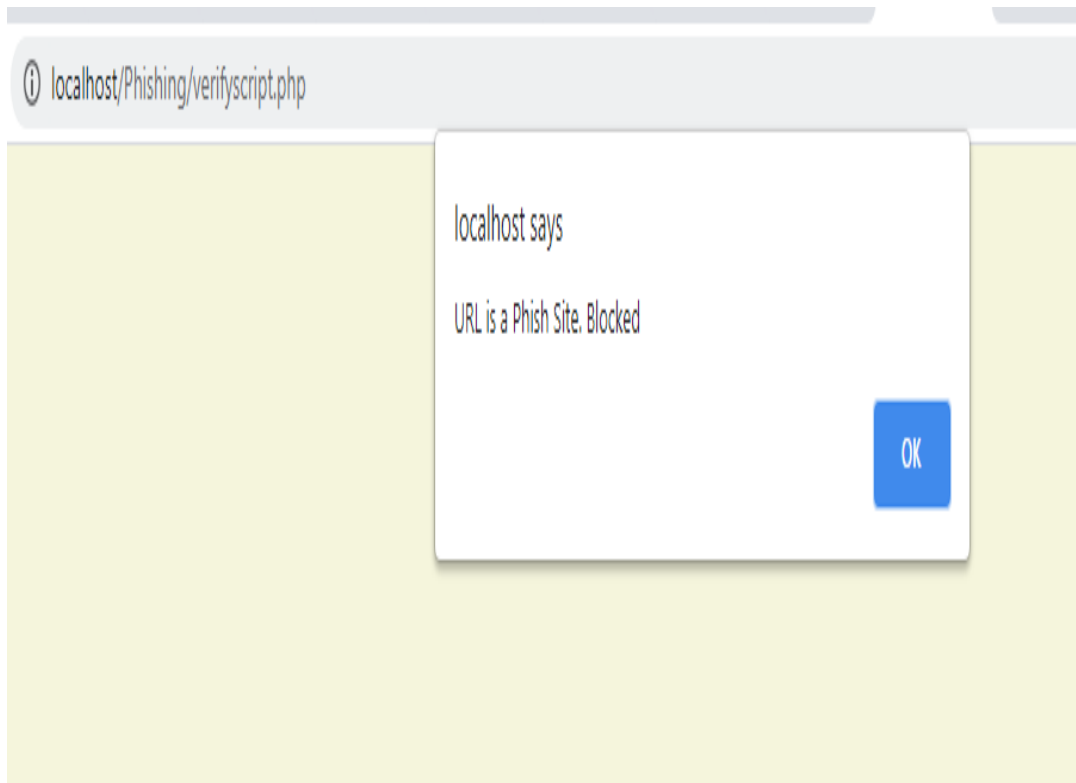


Fig. 9: Phish Detection Result

No admin is required in this system as the process of validation is automated and carried out seamlessly.

The most important component of the proposed system is the URL verification form.

If a URL pops on while using the banking application, before any further action is performed (i.e. either a click or ignore) the URL is copied and inputted into the verification prompt. This prompt is directly linked to the database containing the already classified URLs (as either fake or real URLs) using SVM algorithm. On clicking the verification button, the system scans through the database to see which class the URL belongs to. If the URL is found under the fake URL class it is blocked automatically and a phishing site alert is displayed.

Else the URL is approved and the user can have access if they want to.

This is the best way to handle phishing attack irrespective of the platform it pops up on (emails, social media etc.).

Table 1: Performance Parameters

S/ N	Parameter	Existing Model	Proposed Model	Parameter
1	Time Complexity (TC)	7 Minutes	3 Minute	Time Complexity (TC)
2	Bench Mark (BM)	11	13	Bench Mark (BM)

3	Cross Platform Adaptability (CPA)	5	20	Cross Platform Adaptability (CPA)
4	Model Efficiency (ME)	9	15	Model Efficiency (ME)
5	Intrusion Detection Rate (IDR)	18	30	Intrusion Detection Rate (IDR)
6	Identity Integrity Ratio (IIR)	0.5	1.0	Identity Integrity Ratio (IIR)
7	Cost Benefit Analysis (CBA)	10	15	Cost Benefit Analysis (CBA)
8	Risk Assessment (RA)	20	7	Risk Assessment (RA)

$$\text{Total Performance Score} = \frac{\sum \text{Parameters}}{\text{Error Rate}}$$

$$\text{Existing System} = \frac{7+11+5+9+18+0.5+10+20}{1.442} = 55.9$$

$$\text{Proposed System} = \frac{3+13+20+15+30+1.0+15+7}{1.2} = 86.6\%$$

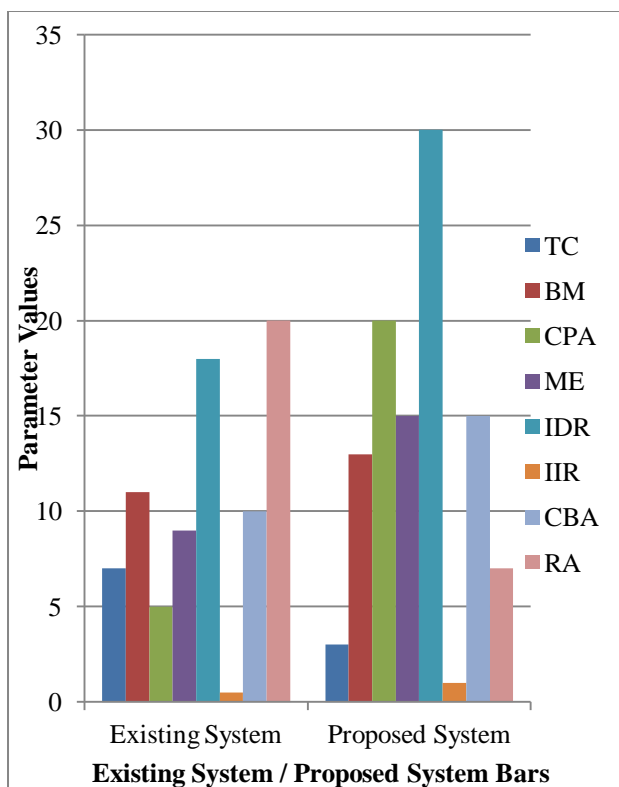


Fig. 10: Performance Evaluation Chart

The system was evaluated using parameters as shown in Table 1. The parameters are used to evaluate the performance of the existing system and the proposed system. From the evaluation of the system, the total performance of the proposed system was

gotten as 86.6%, while the existing system performance was gotten to be 55.9%. Time complexity and model efficiency of the proposed system was higher than that of the existing system.

7. CONCLUSION

This study has provided a solution to the problem of phishing on the banking application platforms. An enhanced model using associative classification was implemented to guarantee secure digital transactions. Several systems and models were surveyed to gain a good understanding of the problem and the solutions in place. Then we proposed a new model to tackle the drawbacks of the existing models. To test this model PHP and MySQL was used to test the practicality of the mode. Performance evaluation has proved our system to be better than the other existing systems.

7.1 Contribution To Knowledge

An enhanced model for secure banking and detection of Phishing sites using associative classification has been presented and simulated in this study which is different from other solutions which were previously proposed

7.2 Future Scope

Integrating other characteristics of phishing via other platforms like voice call phishing detection will further curb the rate of its occurrence. Therefore, in the future, we will integrate a hybrid model including Natural Language Processing algorithms to handling phishing detection on both audio and textual platforms.

8. REFERENCES

- [1] M. I. A. Ajlouni, W. Hadi and J. Alwedyan. 2013 Detecting Phishing Websites Using Associative Classification. *Journal of Information Engineering and Applications*. Vol. 3, No.7, pp. 6-10..
- [2] V. P. Ratnaparkhi, S. S. Jambhulkar. 2020 Framework for Detection and Prevention of Phishing Website Using Machine Learning Approach. *Journal of critical Reviews*. Vol. 7, Issue 7, pp. 2108-2124..
- [3] J. Pawan, W. Tian, P. Li, T. Wei and Z. Liang, 2018 Phishing-Alarm: Robust and Efficient Phishing Detection via Page Component Similarity, *IEEE Access* Vol. 5..
- [4] Z. Futai, G. Yuxiang, P. Bei, P. Li and L. Li. 2016 Web Phishing Detection Based on Graph Mining, 2nd IEEE International Conference on Computer and Communications (ICCC).
- [5] G. Armano, S. Marchal and N. Asokan, 2016 Real-Time Client-Side Phishing Prevention Add-on, IEEE 36th International Conference on Distributed Computing Systems (ICDCS).
- [6] A. K. Jain and B. B. Gupta 2017. Phishing Detection: Analysis of Visual Similarity Based Approaches. *Hindawi Security and Communication Networks*. Pp. 1-20..
- [7] M. Aburrous, M.A. Hossain, K. Dahal and F. Thabata. 2009 Modeling Intelligent Phishing Detection System for e-Banking using Fuzzy Data Mining. *International Conference on Cyberworlds*. Pp. 265-272
- [8] A. Kazi, F. M. Mirkar, G. S. Patil, R. R. Kasar. 2017 Detecting E Banking Phishing Websites using Associative Classification. *International Journal of Engineering Technology Science and Research*. Vol. 4, Issue 10. Pp. 261-264..
- [9] K. R. Ramya, K. Priyanka, K. Anusha, C. J. Devi and Y. A. S. Prasad 2011. An Effective Strategy for Identifying

- Phishing Websites using Class-Based Approach. *International Journal of Scientific & Engineering Research*, Vol.2, Issue 12, pp. 1-7, Dec..
- [10] H. Alhamad, T. Alzyadh and M. A. Badawi. 2020 Detecting E-Banking Phishing Website using C4.5 Algorithm. *IJCSNS International Journal of Computer Science and Network Security*, Vol.20 No.11, pp. 46-51.
- [11] R.M. Mohammad, F. Thatbtah and L. McCluskey. 2018 Intelligent Rule based Phishing Websites Clasification. *IET Journal*. Pp. 1-22.
- [12] V. Kamble, D.Khobragade, P. Wasnik, D. Yadav and P. Gaidhane. 2017 Detecting E-banking Phishing Websites using Naïve Bayes Classifier. *International Journal for Research in Emerging Science and Technology*. Pp.94-96. .
- [13] V. Divya and V. Vivitha. 2017 Phishing Websites Detection using Associative Classifiers. *International Journal on recent Researches in Science, Engineering and Technology (IJRRSET)*. Vol. 5, Issue 11. Pp. 30-39.
- [14] S. Wankhede, R. Nikose, S. Domle, S. Asatkar and J. Singh. 2018 Detecting the Phishing Websites using Enhance Secure Algorithm. *International Research Journal of Engineering and Technology (IRJET)*. Vol. 5, Issue 3. Pp. 494-495.
- [15] R. Damodaram and M. L. Valarmathi. 2011 Phishing Websites Detection and Optimization using Particle Swarm Optimization. *International Journal of Computer Science and Security (IJCSS)*. Vol. 5, Issue 5. Pp. 477-490.