iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.53177

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Phishing Attack Detection on Text Messages Using Machine Learning Algorithms

Swaraj Uplenchwar[1], Rutika Shinde[2], Srushti Gunthe[3], Rutuja Gholap[4], Ajinkya Kobal[5]
[1, 2]B.E Students, [3]Assistant Professor, [4, 5]B.E Students, Department of Information Technology Sinhgad Institute of Technology and Science, Narhe, Pune, India

Abstract: ***Phishing is the most common form of social engineering assault that tries to trick or take advantage of computer users. Attackers attempt to learn information about someone or something by doing phishing, particularly on text messages. It is crucial to have an efficient technique for the detection of the same because such text message phishing assaults are constantly changing. This study introduces a text message phishing attack detection system (PADSTM) that focuses on detecting phishing assaults in text messages using machine learning (ML). To identify the phished communications, it employs machine learning (ML) techniques as Naive Bayes' Classifier, Support Vector Classifier, Random Forest Classifier, and K- Nearest Neighbour Algorithm (KNN). For effective phishing attack detection, PADSTM concentrates on the blacklist of URLs and numerous customised keywords in the text messages. According to experimental findings, Random Forest Classifier outperforms other ML approaches in terms of accuracy and F1-score when it comes to identifying phished communications.***
Keywords: Cyber security, Machine Learning (ML), Phishing, Text classification, URL blacklist

I. INTRODUCTION

Phishing is a type of attack in which a perpetrator sends a false message intended to computer or internet users into disclosing sensitive information or allowing malware to be installed on their systems [1]. Passwords, credit card numbers, social security numbers, and other private information are examples of this sensitive information [2]. In this attack, the attacker lures the target recipients into clicking a link, dialing a phone number, or emailing a contact address, which results in the release of the victim's sensitive information to the attackers. The users of these websites and their associated transactions are increasingly the main targets of hacking as the number of eCommerce websites rises dramatically. The first recorded phishing assault occurred on the E-Gold website in June 2001 [3]. Despite the fact that this attack was unsuccessful, it inspired other ones in the future. 90% of data breaches are caused by phishing, according to a report released by Cisco in 2021 [4]. As a result, it poses the greatest threat to information security. Phishing assaults are constantly changing, resulting in both financial losses and an emotional impact on the victims. Therefore, from the viewpoint of the end user, phishing attack detection is crusial.

In order to verify the validity of the messages, this study introduces the PADSTM phishing attack detection system for text messages. Traditional methods have a high false- positive rate and significantly longer detection times when employed to identify phishing attempts [5]. Machine Learning (ML) models, on the other hand, are effective by nature and can therefore be more useful to detect anomalous materials effectively. Therefore, the Naive Bayes' Classifier, Support Vector Classification (SVC), Random Forest Classifier, and K-Nearest Neighbour Algorithm (KNN) ML techniques are used in the proposed system to detect phishing attacks. Customised text message keywords are used as features in the many ML approaches used to detect phishing assaults. In order to improve the detection of phishing attacks, the suggested system also takes into account checking against a list of harmful URLs known as a blacklist of URLs. Based on accuracy and F1-score, phishing attack detection results utilising ML systems are evaluated. As a result, the Random Forest Classifier exhibits the highest level of accuracy and F1-score.

II. RELATED WORK

Attacks like phishing aim to trick users rather than systems. Early in 2019, the APWG [6] recorded 1,238,161 phishing assaults. The authors detail the several iterations of these attacks as well as the various detection methods in [7], including the heuristic-based approach, the blacklist approach, the ML approach, and the image-based approach.

By extracting verb-direct object pairings and comparing them to a topic blacklist of harmful pairs, SEA Hound [8] is a system that recognises email phishing. Lexical analysis is used in [9] by Phish Haven to extract features. [10].

To identify phishing websites, authors in [11] employ three classifiers with feature selection techniques from weka. The fraudulent URL is examined using URL features by the phishing detection technique suggested in [12]. It is recommended by authors in [13] that detection algorithms take into account all potential variations in assault strategies.

©IJRASET: All Rights are Reserved

Authors in [14] suggested numerous approaches for identifying phished URLs. However,these approaches required ongoing work to fend off fresh threats.

In order to make the system dynamic, [15] establishes a supplemental methodology to the current phishing detection approaches. However, the work has not been approved for use in the actual world. A zero-day defence method was proposed by authors in [16] for detecting phished URLs by relying merely on the website address's content.

The model suggested in [17] comprises two levels: first level filters out spam and ham messages, while the second level separates smished messages from spam messages. There is an overhead in terms of storage, processing, and time because this approach detects phished messages in two steps. By examining short, unstructured communications, authors of [18] pinpoint SMS phishing. A method that incorporates SMS classification and domainvalidation is suggested in [19] as a way to filter smished text messages.

The authors of [20] created a method by combining various techniques to improve the accuracy of spam filtering. A method for detecting phishing assaults in emails that combines ML and NLP is suggested in [21]. The authors of [22] talk about the problems they encountered when utilising deep learning to identify phishing assaults.

To In conclusion, the evaluation of related workmentioned above shows that several methods have been suggested for phishing attack detection in text messages. However, these systems only take a relatively small number of terms into account while looking for phishing attempts. Additionally, these methods do not take URL into account as a characteristic for phishing attacks on text messages. These shortcomings in the preceding work are intended tobe fixed by the proposed PADSTM. The URL is regarded as a crucial characteristic, and the URL in text messages is checked against a list of URLs that are prohibited. Thesuggested system uses machine learning methods to classify text messages into benign or malicious ones depending on their contents. Text message customizations are utilised as features by machine learning (ML) systems to more precisely identify phished messages.

III. METHODOOLGY

In this section, we go through PADSTM's functionaloverview and the specifics of its methodology for detecting phishing attacks.

A. Functional Overview

As depicted in Fig. 1, PADSTM accepts the text message as input and then verifies the message's authenticity. As a result, PADSTM provides a response indicating whether the message is phished or not.

As shown in Fig. 1, PADSTM contains three modules which are described below.

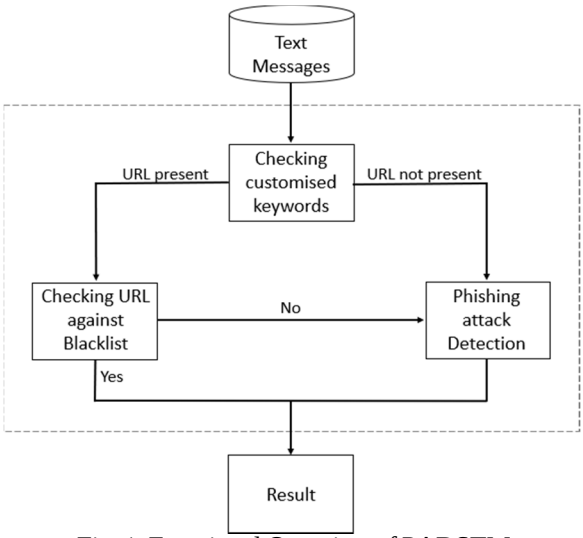


Fig. 1. Functional Overview of PADSTM

1) *Searching for Customized Keywords:* The presence of customised keywords in the text message is examined in thismodule. It is thought of as a set of customised keywords when a particular category of keywords tempts the end usersto take actions in the message that could cause the disclosure of sensitive information. These classifications comprise unique tokens, money signs, morphemes, cellphone numbers, and URLs.

2) *Check against blacklist of URLs:* This module determines whether or not the URLs contained in the text messages have been phished. By comparing the URLs to the blacklist, this verification is carried out. When the text message's URLturns up on a list of prohibited URLs, the system determinesthat the message is a phished one. If the aforementioned criterion is not true, the text message proceeds to the next stage of the system to determine whether it is still a phished message or not.
3) *Phishing Attack Detection:* Using the module for detecting phishing attacks, the text messages with legal URLs aretested one more time for phishing. Additionally, if the URL is absent from the text message, the phishing attack detection module assesses the text message's veracity using the additional categories of customised keywords that were obtained from the message. The outcome is then provided, indicating whether the message is phished or not.

Using the most effective ML algorithm, the text messages are classified in order to identify phishing attacks. The Naive Bayes Classifier, the Support Vector Classifier (SVC), the Random Forest Classifier, and the KNN have all been taken into consideration in this work. Section III-C provides specifics on how to recognise phishing attacks.

B. Searching for customized keywords

The PADSTM system considers specific categories of customised keywords, as was described in Section III-A. Then it looks for these categories' related keywords in the delivered text message.

The types of customised keywords that PADSTM concentrates on are unique tokens, currency symbols, visual morphemes, mobile numbers, and URLs.

By providing various services that users would find useful, the unique tokens attempt to attract their attention. Patterns like "xx.." are examples of visual morphemes, which are frequently employed to conceal private data like account numbers and phone numbers.

Any particular set of keywords in these categories is not constrained by our system PADSTM. A few instances of thecustomised keywords are shown below, however only for demonstration purposes.

Congratulations, hurry, winner, private, unsubscribe, free, reminder, jackpot, prize, cash, awarded, join, claim, and special tokens.

Symbols for money include: $, £, €,, X, X, XX, XXX, morphemes in visual form

Each of the following categories—currency symbols, visual morphemes, mobile numbers, and URLs—may typically only appear once in a malicious text message. Nevertheless, special tokens could appear more than once in a single message. As a result, as explained in the paragraphs that follow, either the entire category or a specific keyword inside the category is thought of as a feature for extracting it from the text message.

One attribute is taken into consideration for the entire category of monetary symbols. This indicates that if any money symbol, such the symbol $, is present in the text message, the feature value for the entire category of currency symbols is set to 1. Setting the feature values for the whole categories of visual morphemes, cellphonenumbers, and URLs uses a similar methodology to that of the category of monetary symbols.

For the special token category, each special token that is specified in the category is regarded as a unique feature. To put it another way, if the special token "congratulations" is included in the text message, the feature value for that specific special token is set to one. In a similar vein, if any more exceptional tokens are discovered in the text message, their feature values are likewise set to 1.

As a result, when removing special tokens from text messages, each one is given equal weight. This leads to a more accurate identification of unique tokens for spotting phished mails.

Using the categories of tailored keywords mentioned above, the comprehensive set of features from the text message is extracted. By utilising ML algorithms, these attributes are also used to categorise text communications asbenign or malicious.

C. Phishing Attack Detection

Python and the Django framework are used to build the PADSTM prototype. The prototype solution makes use of a dataset that contains a list of text messages.

The general strategy for phishing attack detection entails training a dataset of text messages, which is then used to verify the veracity of incoming fresh messages. Based on the many customised keywords that are retrieved from the various categories and used in the training phase of machine learning algorithms for text messages, the messages are categorised as phished or not phished using the ML algorithms. During the testing phase, the ML algorithms use the training set of text messages to categorise each new text message that arrives as phished or not phished. Comparing these labels to the matching actual labels in the collection of text messages allows us to gauge how well the phishing attack detection system performed.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

The best performance ML algorithm for phishing attack detection is used by PADSTM, as was described in Section III-A. For the purpose of detecting phishing attacks, the performance of the ML algorithms Naive Bayes Classifier, SVC, Random Forest Classifier, and KNN is evaluated. Sklearn is a Python package that is used to implement these machine learning methods. Following are few paragraphs that describe how these ML methods are used.

1) *Naive Bayes Classifier:* To determine whether a text message is fraudulent or not, it applies probability theory. To distinguish between the several types of customised keywords in this study, a categorical classifier is used.
2) *Support Vector Classification:* Kernel and random state are the two hyperparameters employed in SVC. This classification approach uses the fit method, which trains the model using the best-fitting hyperplane, asa training set. The data points are divided into phished and non-phished groups using this hyperplane. Since it solves the space complexity issue by just keeping the support vectors during training, the radial basis function (RBF) kernel is chosen in this study. In order to ensure consistency, the random state is set to zero.
3) *Random Forest Classifier:* Two hyperparameters, random state and n estimators, are selected for this model. To ensure consistency in the findings, the random state is set to zero. The ideal number of trees for the ensemble model is 100 n estimators.
4) *K-Nearest Neighbor:* The parameter n neighbors, or the quantity of neighbors, is provided in KNN. The labels of text messages that are the new text message's closest n neighbors are reviewed in order to determine if they have been phished or not. The Euclidean distance is used to determine which n neighbors are the text message's closest.

On the basis of accuracy and F1-score, the aforementioned ML algorithms are contrasted in order to choose the algorithm that is most appropriate. PADSTM uses the machine learning algorithm that detects phishing attacks with the best accuracy and F1-score.

IV. PERFORMANCE EVALUATION

This section elaborates on the dataset which is employed in experimentation, the metrics which are used in evaluating the performance of phishing attack detection and the results achieved.

A. Dataset and data preprocessing

In order to evaluate PADSTM experimentally, the text message dataset [23] was modified. English-language text messages totaling 5572 are part of this dataset. It includes two columns, the first of which lists the text messages and the second of which indicates whether or not each text message was phished. The messages are all changed to lowercase as part of the dataset's preparation. The rate of correctly detecting phished or non-phished texts will be at its highest when using the Random Forest model todetect phishing assaults on a batch of text messages, according to this evidence.

B. Evaluation Metrics

This paper uses accuracy, precision, recall, and F1-score[19] as the metrics for validating the application of MLapproaches in phishing attack detection. These evaluation measures are understood as follows in relation to the proposed PADSTM.

Accuracy is the measure of correctly predicting the incoming new messages as whether they are phished or not phished.

$$Accuracy = \frac{(TP + TN)}{TP + FP + TN + FN}$$

Precision implies the ratio of the number of phished messages that were correctly predicted compared to all other phished messages that were correctly anticipated.

$$Precision = \frac{TP}{TP + FP}$$

Recall the ratio between the number of phished messages that were accurately anticipated and the total number of real phished messages.

$$Recall = \frac{TP}{TP + FN}$$

F1-score is a value that represents the harmonic mean of Precision and Recall.

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

In our case, it is equally crucial for both scenarios—phished messages being correctly identified as such andnon-phished messages being accurately identified as such—to occur. Therefore, we employ the F1-score as a statistic rather than just Precision or just Recall.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

C. Results and Analysis

Based on accuracy and F1-score, ML algorithms employed in phishing attack detection are compared for their performances. 73% of the dataset [23] is used fortraining purposes during experimentation, while 27% is used for testing purposes. The testing data, which consists ofa set of 1504 text messages out of 5572 text messages, is used to calculate the accuracy and F1-score.

Table I lists the accuracy and F1-score figures for the ML algorithms that were discovered through experimentation. Additionally, it provides the precision and recall values that are used to determine the associated F1- score.

Table I. Comparative Results Of Phishing Attack Detection

ML Algorithm	Accuracy	Precision	Recall	F1-score
Naive Bayes	0.9678	0.9954	0.9673	0.9811
SVC	0.9660	0.9938	0.9682	0.9808
Random Forest	0.9690	0.9962	0.9665	0.9811
K-Nearest Neighbor	0.9641	0.9985	0.9616	0.9797

Fig. 2 depicts the graphical results of the accuracy of phishing attack detection, obtained during the experimentation.

The results in Table I and Fig. 2 show that the Random Forest model has the maximum accuracy among the other ML algorithms used. This suggests that the Random Forest model will produce the lowest classification error when used to detect phishing attacks on a batch of text messages, and the rate of properly identifying phished or non-phished texts will be at its greatest.

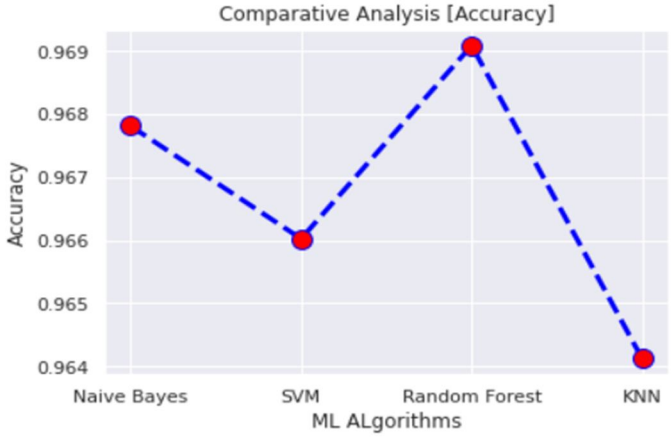


Fig. 2. Accuracy in Phishing Attack Detection

Fig. 3 illustrates the graphical results of the F1-score inphishing attack detection, obtained during the experimentation.

The results in Table I and the Fig. 3 show that RandomForest Classifier has the highest F1-score among the otherML algorithms.

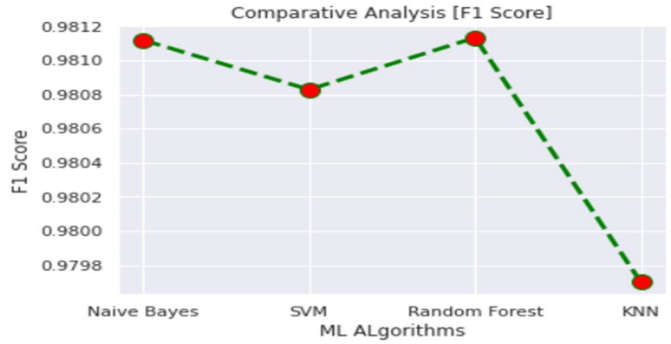


Fig. 3. F1-score in Phishing Attack Detection

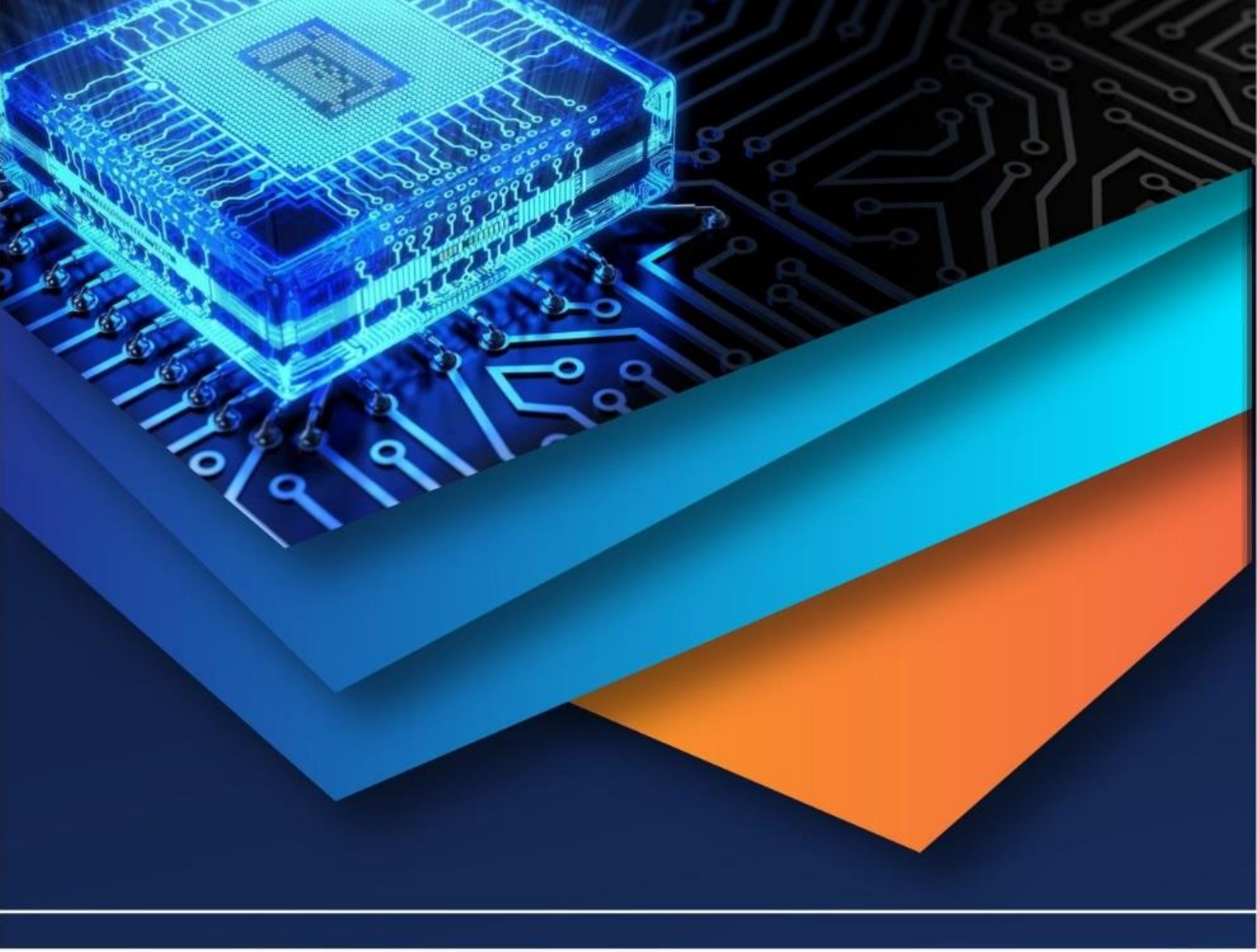
Thus, as the Random Forest model depicts the maximumaccuracy and the F1-score, it is the best suited technique for phishing attack detection in the proposed system.

V. CONCLUSION

A PADSTM technique to identify phishing in text messages has been introduced in this paper. This work's most important addition is its ability to correctly identify phishing using specific text message keywords, while also taking URL verification using a blacklist and ML approaches. The best method for detecting phishing attacks is to use the proposed PADSTM, which involves comparing the text message content to a blacklist of URLs before classifying it. We evaluated how Naive Bayes, SVC, Random Forest, and KNN performed in terms of accuracy and F1-score. Theaccuracy and F1-score of the Random Forest Classifier hold the greatest values, according to experimental data. In light of this, it has been determined that the Random Forest Classifier is the model that will work the best at spotting phishing attempts in text messages. Any text message, including SMS, WhatsApp, and textsfrom social media sites like Twitter or Instagram, can be phished, and PADSTM is capable of seeing it. Thus, while handling various types of text messages, the suggested approach enables users to successfully defend against phishing assaults. Future phishing attack detection methods may take into account text messages produced in languages other than English.

REFERENCES

[1] AL-Otaibi, Abeer F., and Emad S. Alsuwat. "A study on social engineering attacks: phishing attack." International Journal of Recent Advances in Multidisciplinary Research 7, no. 11 (2020): 6374-6380
[2] Ali, Mazurina Mohd, and Nur Farhana Mohd Zaharon. "Phishing—A Cyber Fraud: The Types, Implications and Governance." International Journal of Educational Reform (2022): 10567879221082966
[3] Prince, Md Sirajum Munir, Asib Hasan, and Faisal Muhammad Shah. "A New Ensemble Model for Phishing Detection Based on Hybrid Cumulative Feature Selection." In 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), pp. 7- 12. IEEE, 2021.
[4] Rosengren, Kim. "Contribution of Open-Source Intelligence to Social Engineering Cyberattacks." (2022).
[5] Odeh, Ammar, Ismail Keshta, and Eman Abdelfattah. "PHIBOOST-a novel phishing detection model using Adaptive boosting approach." Jordanian Journal of Computers and Information Technology (JJCIT)7, no. 01 (2021).
[6] Balim, Caner, and Efnan Sora Gunal. "Automatic detection of smishing attacks by machine learning methods." In 2019 1st International Informatics and Software Engineering Conference (UBMYK), pp. 1-3. IEEE, 2019.
[7] Kathrine, G. Jaspher Willsie, Paradise Mercy Praise, A. Amrutha Rose, and Eligious C. Kalaivani. "Variants of phishing attacks and their detection techniques." In 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), pp. 255-259. IEEE, 2019.
[8] Peng, Tianrui, Ian Harris, and Yuki Sawa. "Detecting phishing attacksusing natural language processing and machine learning." In 2018 ieee 12th international conference on semantic computing (icsc), pp. 300-301. IEEE, 2018
[9] Sameen, Maria, Kyunghyun Han, and Seong Oun Hwang. "PhishHaven—an efficient real-time ai phishing URLs detection system." IEEE Access 8 (2020): 83425-83443
[10] Garcés, Ivan Ortiz, Maria Fernada Cazares, and Roberto Omar Andrade. "Detection of phishing attacks with machine learning techniques in cognitive security architecture." In 2019 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 366-370. IEEE, 2019
[11] Mehanović, Dželila, and Jasmin Kevrić. "Phishing Website Detection Using Machine Learning Classifiers Optimized by Feature Selection." Traitement du Signal 37, no. 4 (2020)
[12] Jain, Ankit Kumar, and B. B. Gupta. "PHISH-SAFE: URL features- based phishing detection system using machine learning." In Cyber Security, pp. 467-474. Springer, Singapore, 2018
[13] Sahingoz, Ozgur Koray, Ebubekir Buber, Onder Demir, and Banu Diri. "Machine learning based phishing detection from URLs." ExpertSystems with Applications 117 (2019): 345-357.
[14] Pradeepa, G., and R. Devi. "Review of Malicious URL Detection Using Machine Learning." In Soft Computing for Security Applications, pp. 97-105. Springer, Singapore, 2022.
[15] Chatterjee, Moitrayee, and Akbar-Siami Namin. "Detecting phishing websites through deep reinforcement learning." In 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), vol. 2, pp. 227-232. IEEE, 2019
[16] Wei, Wei, Qiao Ke, Jakub Nowak, Marcin Korytkowski, Rafał Scherer, and Marcin Woźniak. "Accurate and fast URL phishing detector: a convolutional neural network approach." ComputerNetworks 178 (2020): 107275.
[17] Jain, Ankit Kumar, Sumit Kumar Yadav, and Neelam Choudhary. "A novel approach to detect spam and smishing SMS using machine learning techniques." International Journal of E-Services and Mobile Applications (IJESMA) 12, no. 1 (2020): 21-38.
[18] Boukari, Badr Eddine, Akshaya Ravi, and Mounira Msahli. "Machine learning detection for smishing frauds." In 2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC), pp. 1-2. IEEE, 2021.
[19] Mishra, Sandhya, and Devpriya Soni. "DSmishSMS-A System to Detect Smishing SMS." Neural Computing and Applications (2021): 1-18.
[20] Baaqeel, Hind, and Rachid Zagrouba. "Hybrid SMS Spam Filtering System Using Machine Learning Techniques." In 2020 21st International Arab Conference on Information Technology (ACIT), pp. 1-8. IEEE, 2020.
[21] Kumar, Abhishek, Jyotir Moy Chatterjee, and Vicente García Díaz."A novel hybrid approach of svm combined with nlp and probabilisticneural network for email phishing." International Journal of Electrical and Computer Engineering 10, no. 1 (2020): 486.
[22] Do, Nguyet Quang, Ali Selamat, Ondrej Krejcar, Enrique Herrera- Viedma, and Hamido Fujita. "Deep Learning for Phishing Detection: Taxonomy, Current Challenges and Future Directions." IEEE Access (2022)
[23] https://www.kaggle.com/galactus007/sms-smishing-collection-data- set[Accessed on 15th June, 2022]
[24] https://www.kaggle.com/taruntiwarihp/phishing-site- urls?select=phishing_site_urls.csv [Accessed on 15th June, 2022]
[25] https://www.kaggle.com/sid321axn/malicious-urls-dataset [Accessed on 15th June, 2022]



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)