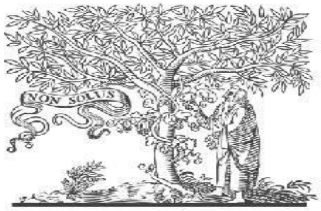
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER SSRN

IJIEMR Transactions, online available on 10th Apr 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 04

10.48047/IJIEMR/V12/ISSUE 04/88

Title **CLASSIFICATION OF PHISHING WEBSITES USING MULTI-LAYER PERCEPTRON**

Volume 12, ISSUE 04, Pages: 721-727

Paper Authors

Ms. G. Sireesha, V. Veda Naga Vyshnavi, S. Sravani, T. Leela Srivaishnavi Devi, Sk.Shehanaz

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

CLASSIFICATION OF PHISHING WEBSITES USING MULTI-LAYER PERCEPTRON

Ms. G. Sireesha[1], Assistant Professsor ,Department of IT,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

V. Veda Naga Vyshnavi[2], **S. Sravani**[3], **T. Leela Srivaishnavi Devi**[4], **Sk.Shehanaz**[5]
[2345]UG Students, Department of IT,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
vyshnavivecha@gmail.com, sravanisomineni2002@gmail.com,
leelasrivaishnavi2001@gmail.com, shaikshannu2932@gmail.com

Abstract

Websites in today's world serve several functions.Web security worries are on the rise along with the daily increase in internet users. Cyber attacks on individuals are becoming more and more commonplace. Phishing is among the often occurring web attacks. Phishing is a social engineering attack method that is frequently employed to acquire user-sensitive data, such as login credentials, credit and debit card information, and so forth. Phishing websites mimic the name and design of a legitimate website. Commonly known as a fake website, it tries to trick visitors into giving up their identities. Maximizing user protection against phishing websites was one of the main objectives in developing these models. With clever phishing detection management techniques, designers can contribute to the achievement of this objective. In this study, we describe an unique method for detecting phishing websites on the client-side using a machine learning algorithm. We use the extraction framework rule in this system paper to extract a website's attributes from just its URL. The proposed method makes use of a dataset containing 30 different URL attributes, which the same Multilayer Perceptron Classification machine learning model would make use of to evaluate the legitimacy of the website. 11,055 tuples make up the dataset used to train the model. The proposed approach results in a strong performance on the 80:20 split ratio.

Keywords: Phishing, Cyber Security, Machine Learning, Multi-Layer Perceptron(MLP), Fraud Detection, Neural Network, Sensitivity Analysis

Introduction

Phishing is a type of cyber-attack that uses phoney websites or emails to persuade users to divulge private data, including usernames, passwords, and credit card numbers. The potential of phishing attacks has grown significantly in importance for both individuals and corporations as more people and businesses use the internet for financial transactions and online shopping. To counter the threat posed by phishing, researchers and security professionals have created a number of methods for spotting and categorising phishing websites. One such technique is the

application of multi-layer perceptron (MLP) neural networks and machine learning algorithms. For supervised learning tasks like classification, an artificial neural network called a multi-layer perceptron is used [1] . A simple calculation is carried out by each node in the multi-layer MLP network using its inputs and a set of weights. Before the final output is produced, the nodes in one layer's output are fed into the next layer.A dataset of well-known phishing and authentic websites is used to train the MLP network for the classification of phishing websites. The network gains knowledge on the patterns and characteristics that set apart phishing websites from trustworthy ones. After being educated, the network can be used to categorise new websites as legitimate or phishing based on their properties. A potential method that has demonstrated good accuracy rates in prior studies is the classification of phishing websites using MLP. By constructing and evaluating an MLP-based model on a real-world dataset, this study intends to investigate the efficacy of MLP for classifying phishing websites. The project will comprise pre-processing the data, feature extraction, model training and evaluation, and finally the creation of a web application for real-time categorization of phishing websites.

2. Literature Survey

MLP has been used in several studies to categorize phishing websites. For instance, a dataset of legitimate and phishing websites was used by Li et al study in 2019 to train an MLP model. The algorithm successfully identified phishing with a test set accuracy of 97.6%.In a research by Sharma et al. (2020) [2], an MLP model was trained using a dataset of phishing websites. In comparison to other machine learning methods like logistic regression and decision trees, the model outperformed them with an accuracy of 98.3%.The security of websites has been the subject of numerous research papers; some of them have altered routing security (Salehi, Boukerche, and Darehshoorzadeh, 2016); others have worked on intrusion detection, intrusion prevention, and smart grid security (Delgado-Gomes, Martins, Lima & Borza, 2015).First introduced in, the machine learning-based PLIFER model developed by the writers (Abdelhamid, Thabtah & Abdel-jaber, 2017). How long has the URL name been in use? is necessary for this approach. In addition to the phishing website, ten other features that were retrieved are used in the Random Forests model (RF) to recognize it. Using this algorithm, 96% of instances were found.In order to recognize scam websites, the authors of (Zhu, Ju, Chen, Liu & Fang, 2020) employed an artificial neural network. To determine whether a website is a phishing scam, the suggested work used 2 neurons as the output, 17 neurons as the input for 17 features, and 1 hidden layer level. The data gathering resulted in the creation of an experimental set and a train set. The proposed model yielded a value of accuracy of 92.48%.In (Pandey, Gill,

Nadendla, & Thaseen, 2018) [3], the authors' work was concentrated on establishing a consensus regarding the characteristics that are used to identify phishing on websites. In order to detect incursion on web pages using three common data sets, the authors used the Fuzzy Rough Set (FRS) theory.In (Al-Sarem et al., 2021), the authors identified the components that were most effective at identifying website scams and provided two brand-new machine learning-based methods for selection or detection. Both techniques use classifiers called AdaBoost and LightGBM.Sharma and Gupta's (2021) ensemble of MLP models was used in a different research to classify phishing websites. The ensemble was made up of various MLP models that were each trained using a distinct subset of the dataset, and the results were then combined by means of a voting system. The ensemble's 99.4% accuracy on the test set proved how resistant it was to changes in the raw data.It has been demonstrated that the hybrid classifier created by combining these two techniques can identify web phishing attacks more accurately than a single classifier. This program successfully recognized phishing emails in 96% of instances. Through the use of annotated data sets, classification systems can recognize phishing. shows the Hybrid Set Of Features (HEFS) model, which is a recommended software collection model that uses machine learning to detect phishing websites. The basic feature collection is extracted using a method known as the cumulative distribution gradient.The results of the trial show 96.26 percent accurate meta-learners who are extremely effective. The feature extraction method is used by the majority of current machine learning methods and is very effective at detecting phishing. It is possible to obtain more than 200 traits, claims (Khalid, Khalil, & Nasreen, 2014) [4]. A classifier gets bigger when there are more characteristics, but this can cause problems with overfitting.

3. Problem Identification

The problem identification for classification of phishing websites using MLP can be summarized as follows:

1.**Phishing is a serious security threat**: Phishing attacks are a frequent strategy used by cybercriminals to acquire sensitive data from consumers. Thus, efficient techniques for identifying and thwarting phishing attempts are required.

2.**Manual methods are not sufficient**: Blacklists and manual inspection are two manual techniques for finding phishing websites [5], although they are not always reliable because they may not be up to date and may miss freshly developed phishing websites.

3.**Need for a robust classification system**: Consideration must be given to data preprocessing, feature selection, model design, training, and testing in order to create a viable classification system employing MLP. The accuracy of the features, model architecture, and data

quality all play a role in the system's performance.

In general, the problem identification for the categorization of phishing websites using MLP entails understanding the potential of machine learning approaches to solve this problem as well as the need for a more accurate and efficient method of identifying phishing websites. The difficulty lies in creating a reliable classification system that can recognise phishing websites with precision while reducing false positives and false negatives**.**

4. Proposed Methodology

This system was developed using Multi Layer Perceptron (MLP) Classifier.

MLP Classifier

Classification tasks are carried out using a Multi-Layer Perceptron (MLP) classifier, a particular kind of neural network. An input layer, one or more hidden layers, and an output layer are only a few of the many layers of interconnected neurons that make up this system.While using the MLP classifier, a set of input characteristics are fed into the network, where they are processed to create a set of output values that represent the anticipated class probabilities. In order to reduce the discrepancy between the true class labels and the predicted class probabilities, the network is tuned during training by varying the weights of the connections between the neurons.The MLP classifier is capable of handling binary and multi-class classification tasks and is proficient at understanding intricate non-linear correlations between the input data and the class labels. The use of regularisation techniques like L1 and L2 regularisation or dropout can be used to prevent overfitting [6] because it may experience it if the number of hidden layers and neurons in each layer is too high.Speech recognition, picture classification, and natural language processing are just a few of the many applications that MLP classifiers are utilised in. The quality of the data and the hyperparameter settings determine how well they operate, albeit they have frequently shown to be successful.

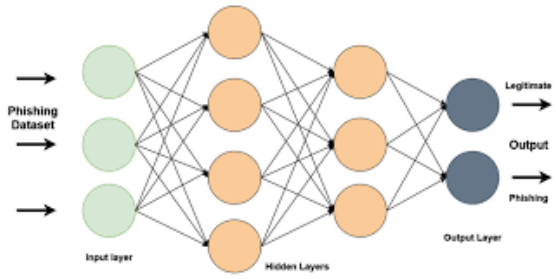


Figure: MLP Architecture

5. Implementation

Implementing a phishing website classification system using MLP involves several steps:

1.**Data Collection**: Make a list of websites that includes both trustworthy and phishing sites. With a total of 11,055 instances and 30 features per instance, the Phishing Websites Data set is made up of phishing websites. The characteristics in the dataset are a mixture of numerical and qualitative variables, including the URL's length, the existence of particular keywords, and the use of special characters. Either a phishing website or a legitimate website is designated for each case.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

#	Attributes		
1	Index	17	Server Form Handler
2	Using IP	18	Info Email
3	Long URL	19	Abnormal URL
4	Short URL	20	Website Forwarding
5	Symbol@	21	Status Bar Cust
6	Redirecting	21	Disable Right Click
7	Prefix Suffix	22	Using Popup Window
8	Sub Domains	23	Iframe Redirection
9	HTTPS	24	Age of Domain
10	Domain Reg Len	25	DNS Recording
11	Favicon	26	Website Traffic
12	Non Std Port	27	PageRank
13	HTTPS Domain URL	28	Google Index
14	Request URL	29	Links Pointing To Page
15	Anchor URL	30	Statistical Report
16	Links In Script tags	31	class

Figure: Dataset Details

2.**Data Preprocessing**: The data is prepped for the MLP by being cleaned and formatted. The data may also need to be normalised and duplicates may need to be handled.

3.**Feature Extraction**: Draw out from the data pertinent properties that can be utilised as MLP inputs. Indicators for phishing website classification include the length of the URL, the age of the domain, the usage of subdomains, and the availability of HTTPS [7].

4.**Train/Test Split**: Divide the preprocessed dataset into a training set and a testing set.

5.**Model Building**: Create an MLP model by utilising a suitable library or framework, such as TensorFlow or Keras. An output layer with a softmax activation function should be included in the MLP along with an input layer and one or more hidden layers [8].

6.**Model Training**: Create the MLP model using the training dataset and a suitable optimization approach, such as stochastic gradient descent. Follow the model's progress on the training set, and make any necessary hyperparameter adjustments.

7.**Model Evaluation**: Analyze the trained MLP model's performance using the testing dataset. To evaluate how well the model works at spotting phishing websites, compute the accuracy, precision, recall, and F1 score.

8.**Model Deployment**: It is possible to deploy the MLP model for usage in practical applications after it has been trained and assessed. This can entail adding the model to a web browser or other piece of software to assist users in spotting and avoiding phishing websites.

Ultimately, a phishing website categorization system utilising MLP necessitates a mix of data collection, pre-processing, feature extraction, model building, training, evaluation, and deployment stages. The accuracy of the data and the MLP model's performance in identifying phishing websites are both critical to the system's success.

6. Results & Conclusions

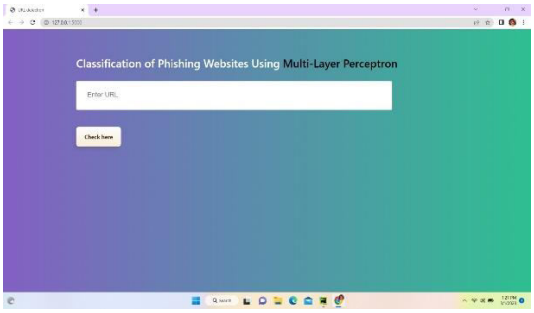


Figure: Overview of homepage of application

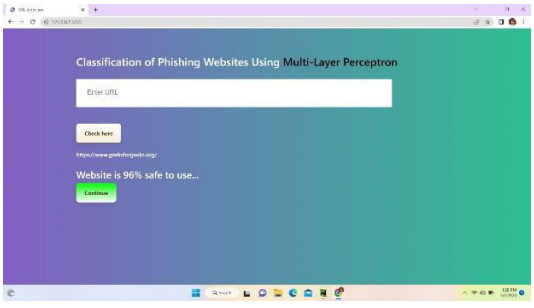


Figure: Incase of Safe URL

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

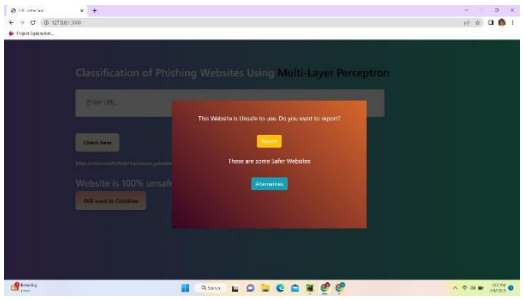


Figure: Incase of Unsafe URL

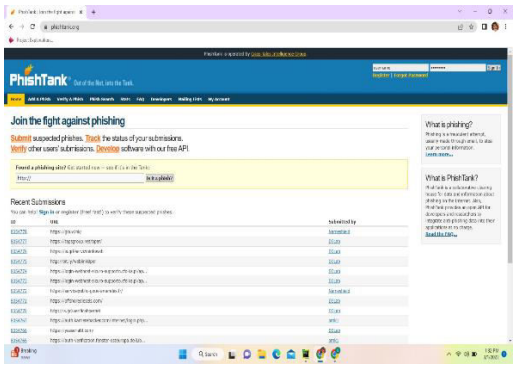


Figure: When user selects Report Option

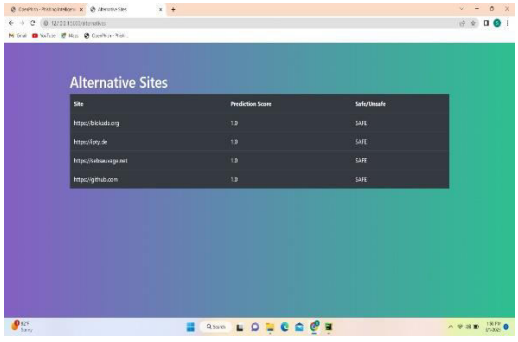


Figure: When user selects Alternatives Option

In conclusion, cybersecurity research is moving in the right direction with the usage of Multi-Layer Perceptron (MLP) for phishing website classification. MLP has demonstrated excellent rates of accuracy in identifying phishing websites, and its capacity to learn from huge datasets and generalise well to new data makes it an effective tool in the struggle against phishing attempts. Although MLP is a potent technique, it shouldn't be used as the only defence against phishing assaults, it is crucial to remember. To supplement the usage of MLP and build a more complete protection against phishing assaults, additional measures like user education and awareness should also be adopted. In general, the use of MLP for phishing website classification is a great addition to the realm of cybersecurity and offers a crucial layer of defence against phishing attempts for both individuals and enterprises.

7. Limitations & Future Scope

Although MLP has demonstrated promising results in the classification of phishing websites, this method has a number of drawbacks:

1. **Low generalisation**: MLP models are prone to overfitting [9], which means they may work well on training data but may not translate well to new, untested data. The model may be less capable of spotting phishing attacks that had not yet been observed as a result.

2. Evolving tactics :It can be challenging to train an MLP model that can precisely recognise all sorts of phishing attempts since attackers who use phishing are continuously coming up with new strategies to avoid detection [10]. To adapt to shifting strategies, the model might need to be periodically retrained.

Some of the **future scope** for this area includes:

1.Ensemble techniques: MLP models can be made to perform better and generalise more widely by using ensemble approaches like bagging, boosting, and stacking. With the aid of these methods, several MLP models that were trained

using various hyperparameters or on various subsets of the data can be combined.

2.Real-time detection: To avoid losses in money and reputation, real-time identification of phishing assaults is essential. The development of quick and effective MLP models with real-time phishing detection capabilities can be the focus of future study.

Overall, there are a number of opportunities to enhance the robustness, accuracy, and speed of the approach, making the future potential for classifying phishing websites using MLP quite encouraging.

References

[1] Riedmiller, M. (1994). Advanced supervised learning in multi-layer perceptrons—from backpropagation to adaptive learning algorithms. Computer Standards & Interfaces, 16(3), 265-278.

[2] Saha, I., Sarma, D., Chakma, R. J., Alam, M. N., Sultana, A., & Hossain, S. (2020, August). Phishing attacks detection using deep learning approach. In 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 1180-1185). IEEE.

[3] Odeh, A., Alhaol, I. A., & Abushakra, A. PHISHING WEBSITE DETECTION USING MULTILAYER PERCEPTRON.

[4] Akram, T., Khan, M. A., Sharif, M., & Yasmin, M. (2018). Skin lesion segmentation and recognition using multichannel saliency estimation and M-SVM on selected serially fused features. Journal of Ambient Intelligence and Humanized Computing, 1-20.

[5] Rao, R. S., & Pais, A. R. (2017). An enhanced blacklist method to detect phishing websites. In Information Systems Security: 13th International Conference, ICISS 2017, Mumbai, India, December 16-20, 2017, Proceedings 13 (pp. 323-333). Springer International Publishing.

[6] Phaisangittisagul, E. (2016, January). An analysis of the regularization between L2 and dropout in single hidden layer neural network. In 2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS) (pp. 174-179). IEEE.

[7] Jeeva, S. C., & Rajsingh, E. B. (2016). Intelligent phishing url detection using association rule mining. Human-centric Computing and Information Sciences, 6(1), 1-19.

[8] Sharma, S., Sharma, S., & Athaiya, A. (2017). Activation functions in neural networks. Towards Data Sci, 6(12), 310-316.

[9] Lawrence, S., & Giles, C. L. (2000, July). Overfitting and neural networks: conjugate gradient and backpropagation. In Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium (Vol. 1, pp. 114-119). IEEE.

[10] Ahmed, N., Amin, R., Aldabbas, H., Koundal, D., Alouffi, B., & Shah, T. (2022). Machine learning techniques for spam detection in email and IoT platforms: analysis and research challenges. Security and Communication Networks, 2022, 1-19.