e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

MALICIOUS URL CLASSIFICATION USING MLP

MR. K. Sudhaakar*1, T. Bharath Reddy*2, S. Srikanth Royal*3,
Abhishek Renwa*4, A. Kiran*5
*1,2,3,4.5Bharath Institute Of Higher Education And Research, Chennai, Tamil Nadu, India.

DOI : https://www.doi.org/10.56726/IRJMETS52705

ABSTRACT

Phishing websites have proven to be a major security concern. Several cyber-attacks risk the confidentiality, integrity, and availability of company and consumer data, and phishing is the beginning point for many of them. Many researchers have spent decades creating unique approaches to automatically detect phishing websites. While cutting-edge solutions can deliver better results, they need a lot of manual feature engineering and aren't good at identifying new phishing attacks. As a result, finding strategies that can automatically detect phishing websites and quickly manage zero-day phishing attempts is an open challenge in this field. The web page in the URL which hosts that contains a wealth of data that can be used to determine the web server's maliciousness. Machine Learning is an effective method for detecting phishing.

I. INTRODUCTION

Phishing is the most unsafe criminal exercises in cyber space. Since most of the users go online to access the services provided by government and financial institutions, there has been a significant increase in phishing attacks for the past few years. Phishers started to earn money and they are doing this as a successful business. The reason for creating these websites is to get private data from users like account numbers, login id, passwords of debit and credit card, etc. Moreover, attackers ask security questions to answer to posing as a high-level security measure providing to users. When users respond to those questions, they get easily trapped into phishing attacks.

II. METHODOLOGY

Webpage: The user accesses a webpage that provides an interface to input a URL for classification. Enter URL: The user pastes or types the URL they want to classify in the input field provided on the webpage. Send URL to Flask Backend: The webpage sends the input URL to the Flask backend, which is responsible for processing and classifying the URL.

III. MODELING AND ANALYSIS

The data is split into 8000 training samples and 2000 testing samples, before the ML model is trained. It is evident from the dataset that this is a supervised machine learning problem. Classification and regression are the two main types of supervised machine learning issues. Because the input URL is classed as legitimate (0) or phishing (0), this data set has a classification problem.

The following supervised machine learning models were examined for this project's dataset training

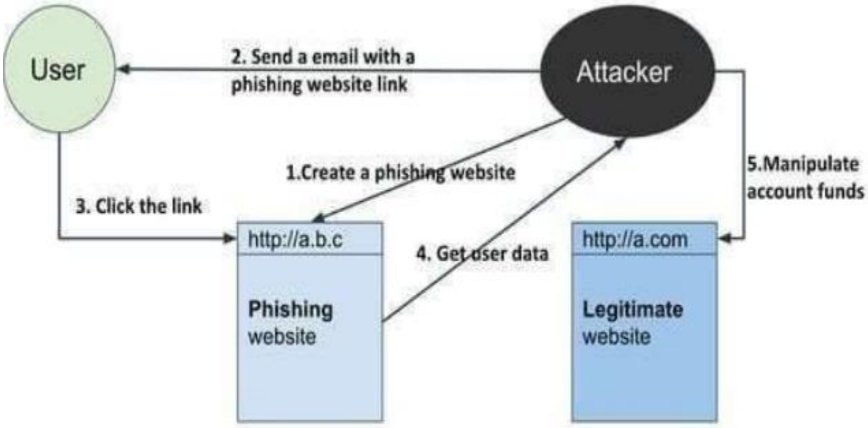
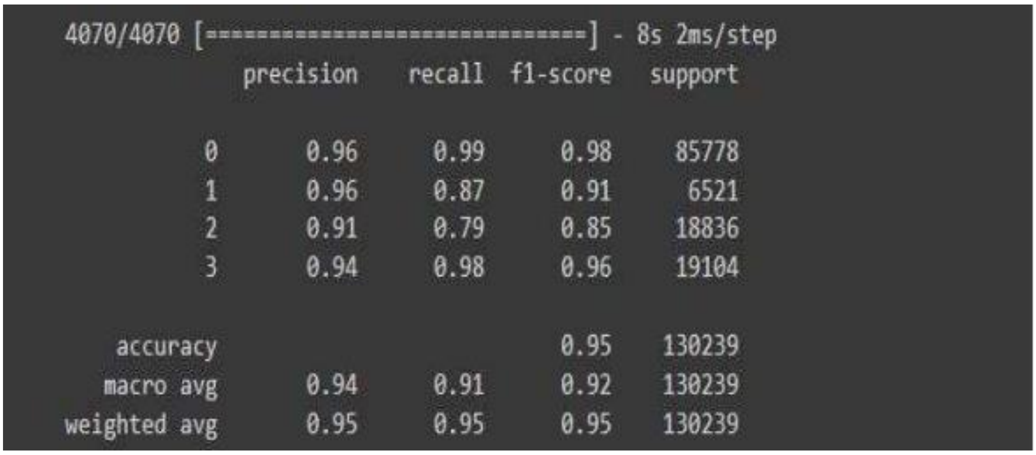


Figure 1: Flow Diagram

IV. RESULTS AND DISCUSSION

The below fig 8.1 explains the completion of the testing, the results revealed that the model had an accuracy of 0.947 and a test loss of 0.1615. According to these measures, the performance of the model with respect to the classification job seems to be satisfactory. A test loss of 0.1615 indicates that the model is capable of properly predicting the right class for an extremely high percentage of the test occurrences. It would seem that the model is picking up on the underlying patterns in the data because of its ability to generalize successfully to new cases.



```
4070/4070 [==============================] - 8s 2ms/step
              precision    recall  f1-score   support

           0       0.96      0.99      0.98     85778
           1       0.96      0.87      0.91      6521
           2       0.91      0.79      0.85     18836
           3       0.94      0.98      0.96     19104

    accuracy                           0.95    130239
   macro avg       0.94      0.91      0.92    130239
weighted avg       0.95      0.95      0.95    130239
```

Figure 2:

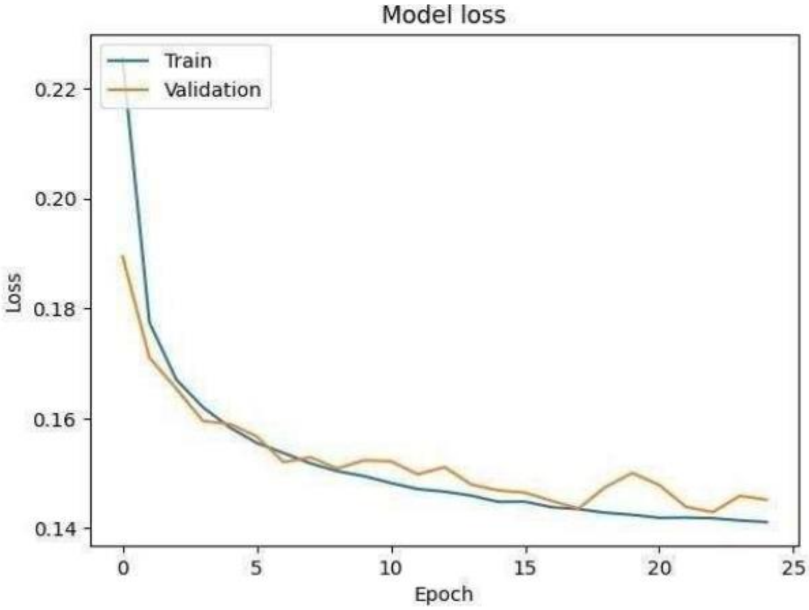


Figure 3: Name of Graph

V. CONCLUSION

This survey presented various algorithms and approaches to detect phishing websites by several researchers in Machine Learning. On reviewing the papers, we came to a conclusion that most of the work done by using familiar machine learning algorithms like K-Nearest Neighbors.

ACKNOWLEDGEMENTS

We express our sincere thanks to our beloved Honorable Chairman Dr S. Jagathrakshakan, M.P., for continuous and constant encouragement in all academic activities. We express our deepest gratitude to our beloved President Dr. J. Sundeep Aanand, and Managing Director Dr. E. Swetha Sundeep Aanand for providingus the necessary facilities for the completion of our project. We take great pleasure in expressing our sincere thanks to Dr. K. VijayaBaskar Raju Pro Chancellor, Dr. M. Sundararajan Vice Chancellor (i/c), Dr. S. Bhuminathan Registrar and Dr. R. Hariprakash Additional Registrar, Dr. M. Sundararaj Dean Academics for moldings our thoughts to complete our project. We thank our dean,

VI. REFERENCES

[1] ‘APWG | Unifying The Global Response To Cybercrime’ (n.d.) available: https://apwg.org/

[2] 14 Types of Phishing Attacks That IT Administrators Should Watch For [online] (2021) https://www.blog.syscloud.com,available:https://www.blog.syscloud.comtypes-ofphishing/

[3] Lakshmanarao, A., Rao, P.S.P., Krishna, M.M.B. (2021) ‘Phishing website detection using novel machine learning fusion approach’, in 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Presented at the 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), 1164–1169

[4] H. Chapla, R. Kotak and M. Joiser, "A Machine Learning Approach for URL Based Web Phishing Using Fuzzy Logic as Classifier", 2019 International Conference on Communication and Electronics Systems (ICCES), pp. 383-388, 2019, July

[5] Vaishnavi, D., Suwetha, S., Jinila, Y.B., Subhashini, R., Shyry, S.P. (2021) ‘A Comparative Analysis of Machine Learning Algorithms on Malicious URL Prediction’, in 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Presented at the 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 1398–1402

[6] Microsoft, Microsoft Consumer safety report. https://news.microsoft.com/ensg/2014/02/11/microsoft-consumersafety-index-reveals-impact-of-poor-online-safetybehaviours-in-Singapore/sm.001xdu50tlxsej410r11kqvksu4nz.

[7] Internal Revenue Service, IRS E-mail Schemes. Available at https://www.irs.gov/uac/newsroom/consumers-warnedof-new-surge-in-irs-email-schemesduring-2016-tax-season-tax-industry-also-targeted.

[8] Abu-Nimeh, S., Nappa, D., Wang, X., Nair, S. (2007), A comparison of machine learning techniques for phishing detection. Proceedings of the Anti-phishing Working Groups 2nd Annual ECrime Researchers Summit on - ECrime ’07