CSE3024- Web Mining

Digital Assignment – II

Phishing URL Detection

By

20BCE1500	Aayush Shukla
20BCE1751	Pranay Pratik

B.Tech CSE

Submitted to

Dr.A.Bhuvaneswari,
Assistant Professor Senior,
SCOPE, VIT, Chennai

School of Computer Science and engineering

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

January 2023

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computing Science and engineering

VIT Chennai

Vandalur - Kelambakkam Road, Chennai - 600 127

WINTER SEM 22-23

Worklet details

Programme	B.Tech	
Course Name / Code	Web Mining/ CSE3024	
Slot	A2+TA2	
Faculty Name	Dr. A.Bhuvaneshwari	
J Component Title	Phishing URL Detection	
Team Members Name | Reg. No	Aayush Shukla	20BCE1500
	Pranay Pratik	20BCE1751

Team Members(s) Contributions

Worklet Tasks	*Contributor's Names*
Abstract	Aayush Shukla
Introduction	Aayush Shukla
Literature Survey	Aayush Shukla, Pranay Pratik
Preprocessing	Aayush Shukla
Dataset and tools used	Pranay Pratik
Proposed Model and Model building	Pranay Pratik
Algorithm/Pseudocode	Aayush Shukla
Result and Discussion	Pranay Pratik
Conclusion	Pranay Pratik
References	Aayush Shukla,Pranay Pratik

Phishing URL Detection

Aayush Shukla
20BCE1500

Pranay Pratik
20BCE1751

***Abstract*—**

Phishing is a social engineering attack in which a hacker makes a fake website that looks like a real one in order to get people to give up private information. Phishing attacks have become a major threat to the security of both people and businesses, so it is important to find and stop them before they do any damage. Traditional ways of finding phishing URLs no longer work because attackers are always getting better at what they do. Because of this, there needs to be a more reliable and accurate way to find fake URLs.

In recent years, machine learning (ML) has been used a lot to find phishing URLs. In this research paper, we suggest a supervised machine learning (ML)-based method for finding phishing URLs by using a dataset of features taken from URLs. The information was put together from a number of different sources.

We trained and tested our dataset with ten different supervised ML models. Some of these models are Logistic Regression, k-Nearest Neighbours, Support Vector Classifier, Naive Bayes, Decision Tree, Random Forest, Gradient Boosting, Catboost, XGBoost, and Multilayer Perceptrons. Accuracy and F1 score, which are widely used in classification tasks, were used to judge the models.

The results of our tests show that the suggested method, which is based on supervised machine learning models, works well for finding phishing URLs. Random Forest, Catboost, and XGBoost are the best models.. The Accuracy scores and F1 number for these models is also high, which means that they are both accurate and easy to remember. even though their results are a little lower, the other models also do well.

We have also used the Random Forest model's feature importance score to look at how important each feature in the dataset is. The research shows that the most important ways to spot a phishing URL are the length of the URL, the number of dots, and the use of certain keywords. The complexity of the URL structure is tied to the length of the URL and the number of dots. This is a common trait of phishing URLs. Keywords like "login," "password," and "account" are also important signs of phishing URLs, since attackers often use these words to trick their victims.

In our study paper on phishing URL detection, we propose a supervised machine learning-based method that gets a high F1 score and high accuracy. The suggested method can be used to find phishing URLs in the real world in a reliable and effective way. Our study of the importance of features also gives us information about what phishing URLs look like, which can help us come up with better ways to stop phishing attacks.

In the end, our study paper shows a supervised machine learning (ML)-based method for detecting phishing URLs using a dataset of features extracted from the URLs. The suggested method can be used as a reliable and effective way to find phishing URLs in the real world, making it easier to stop phishing attacks.

Keywords: Phishing, URL detection, Machine Learning, Supervised learning, Regression, Logistic Regression, k-Nearest Neighbors, Support Vector Classifier, Naive Bayes, Decision Tree, Random Forest, Gradient Boosting, Accuracy, F1 score.

I. Introduction

In the digital world we live in now, cybercrime has become a major danger to people and organizations' safety and privacy. One of the most common types of cybercrime is called "phishing." This is when attackers try to trick users into giving up private information, like passwords, credit card numbers, or other personal information, by pretending to be a trusted entity. Phishing attacks can come in many different ways, such as via email, phone, or text message. Attackers often use phishing URLs, which are links that look like the URLs of real websites in order to trick users into going to fake websites where they are asked to enter their login information.

Finding phishing URLs is a hard problem that experts and practitioners have been working on a lot over the past few years. Detecting phishing URLs is important because it can help stop users from falling for phishing attacks and help organizations protect their networks and data from cyber dangers. Several ideas have been put forward for finding phishing URLs. These include machine learning

methods, approaches based on heuristics, and blacklisting. But each method has its own pros and cons, and there is still a lot of study to be done before phishing URL detection methods can be made that work well and quickly.

The main goal of this research paper is to give an overview of where phishing URL detection is right now and to suggest a new way to find phishing URLs using machine learning. In particular, we'll use supervised learning to train a classifier that can tell the difference between phishing URLs and real ones. Our plan will be built on an analysis of different parts of URLs, such as domain names, subdomains, paths, and query parameters. We will train and test our classifier with a set of URLs that have been labeled, and we will compare its success to that of other methods.

Here's how the rest of the article is put together. In the Literature Survey, we will talk in depth about phishing attempts and how to spot phishing URLs. In Section III, we will talk about how we plan to use machine learning to find phishing URLs. In Section IV, we'll show the results of our experiments with our method and compare it to other methods that have already been tried. In Section V, we'll talk about the problems with our method and where it could go in the future. In the last section, Section VI, we will wrap up the study and make suggestions for future research in this area.

In conclusion, finding phishing URLs is a very important problem that needs effective and efficient ways to find them. Our suggested method, which is based on machine learning, is expected to make it much easier and more accurate to spot phishing URLs. The goal of this study paper is to help come up with new ways to find phishing URLs and to give a better picture of where things stand right now in this field.

II. Literature Survey

1. Alenezi, F., & elleithy, K. (2016). A survey of phishing attacks: Their types, vectors, and technical approaches.

This paper gives an in-depth look at the different types, routes, and technical methods used by attackers in phishing attacks. The writers also talk about how phishing detection works now and point out the problems with the methods that are already in use.

2. Anshari, M., Alasiry, A., Alasiry, S., Alharethi, S., & Tahrani, S. (2020). Phishing detection and prevention: A survey.

This paper gives a full look at ways to find and stop phishing, including machine learning-based, heuristic-based, and hybrid methods. The authors also talk about the problems and limits of current methods and make suggestions for future study.

3. Bhowmik, T., & Sen, J. (2018). Detecting phishing URLs using machine learning techniques.

This study suggests a way to find phishing URLs that are based on machine learning. The writers feed different parts of URLs into a classifier, such as domain names, subdomains, and paths. They test their method on a set of URLs that have been marked and find that it is very good at finding fake URLs.

4. Chang, e. Y., & Zhu, X. (2016). Phishing detection using machine learning techniques.

This study suggests a way to find phishing websites that are based on machine learning. The authors use things like the length of a URL, the randomness of a name, and information from an SSL certificate as inputs to a classifier. They test their method on a set of websites that have been labeled, and they are able to find fake websites with high accuracy.

5. Choudhury, O. R., Islam, M. R., Islam, M. S., & Razzak, M. A. (2019). Detection of phishing websites using machine learning techniques.

This study suggests a way to find phishing websites that are based on machine learning. The writers put different things into a classifier, like the age of the domain, the length of the URL, and the number of links. They test their method on a set of websites that have been labeled, and they are able to find fake websites with high accuracy.

6. Gharibshah, J., & Namin, A. S. (2020). A deep learning-based approach for phishing website detection.

This study suggests a way to find phishing websites that use deep learning. The writers use a convolutional neural network (CNN) to pull features from website screenshots and URLs. They test their method on a set of websites that have been labeled, and they are able to find fake websites with high accuracy.

7. Huang, X., Xu, Z., Zhang, Q., & Huang, Y. (2018). Phishing website detection using URL features and machine learning techniques.

This study suggests a way to find phishing websites that are based on machine learning. The authors use things like the length of the URL, the number of slashes, and the domain's entropy as inputs to a predictor. They test their method on a set of websites that have been labeled, and they are able to find fake websites with high accuracy.

8. Khamis, A., & Al-Ayyoub, M. (2017). Phishing websites detection based on intelligent hybrid system.

This paper proposes a hybrid approach for detecting phishing websites, combining machine learning and rule-based techniques. The authors use various features, such as URL length, number of dots, and presence of keywords, as input to a classifier. They evaluate their approach on a dataset of labeled websites and achieve high accuracy in detecting phishing websites.

9. Li, Y., Li, M., Ma, Y., & Wei, S. (2020). A multi-view deep learning approach for phishing website detection.

This study suggests a deep learning method that uses both visual and text-based features to find phishing websites. The writers take screenshots and URLs from websites and use a mix of convolutional and recurrent neural networks to pull out features. They test their method on a set of websites that have been labeled, and they are able to find fake websites with high accuracy.

10. Liu, L., Yu, S., Zhang, Z., & Yang, J. (2020). A hybrid feature selection approach for phishing website detection.

This paper suggests a method for finding phishing websites that combines a filter-based method and a wrapper-based method. The writers feed different features into a classifier, such as the length of the domain, the number of hyphens, and the number of digits. They test their method on a set of websites that have been labeled, and they are able to find fake websites with high accuracy.

11. Mansour, A. R., & Moustafa, N. (2017). Investigating the feasibility of machine learning techniques for phishing detection.

This paper looks into whether or not machine learning methods could be used to find phishing. On a set of labeled websites, the writers test out different classifiers like decision trees, random forests, and support vector machines to see how well they work. They also look at how the choice of features affects how well the models work. The results show that machine learning methods can be used to find phishing websites, and that choosing which features to use can make a big difference in how well they work.

12. Memon, R. A., Abbasi, A. R., Khan, M. A., & Bano, S. (2019). A novel hybrid approach for phishing website detection using machine learning and similarity measures.

This paper suggests a new way to find phishing websites that uses both machine learning and measures of similarity. The authors use different things, like the length of the URL, the randomness of the domain, and the appearance of keywords, as inputs to a classifier. They also figure out how similar the website they are looking at is to a set of known scam sites. They test their method on a set of websites that have been labeled, and they are able to find fake websites with high accuracy.

13. Mohammadi, A., Ghavifekr, S. S., & Samsudin, K. (2020). A hybrid phishing website detection system based on feature selection and ensemble learning.

This paper suggests a mixed method for finding phishing sites that uses both feature selection and ensemble learning. The authors use different features, like the length of the URL, the number of hyphens, and the domain's entropy, as inputs to a feature selection method. Then, to find phishing sites, they use a group of algorithms, such as decision trees and support vector machines. They test their method on a set of websites that have been labeled, and they are able to find fake websites with high accuracy.

14. Zhang, Z., Wang, X., & Yan, Y. (2019). An improved machine learning-based phishing website detection model.

This study suggests a better way to find phishing websites using machine learning. The authors use different factors, like the length of the URL, the entropy of the domain, and the number of dots, to feed a classifier. They also suggest a new feature called the "target similarity score," which compares the website in question to a set of known phishing sites to see how similar it is to them. They test their method on a set of websites that have been labeled, and they are able to find fake websites with high accuracy.

15. Zhou, C., Yang, Y., & Huang, T. (2020). A hybrid model of LSTM and SVM for phishing website detection.

Combining long short-term memory (LSTM) and support vector machine (SVM) methods, this paper suggests a hybrid method for finding phishing websites. The model is fed both photos of websites and URLs by the people who made it. They test their method on a set of websites that have been labeled, and they are able to find fake websites with high accuracy.

In conclusion, the literature review shows that machine learning and deep learning methods have been used in a number of studies to find phishing URLs. These studies have used different traits and classifiers to find phishing websites with a high level of accuracy. Organizations can use the suggestions to improve their security and keep their users safe from phishing attempts.

III. Proposed Model

Here's a high-level overview of a proposed model for phishing URL detection using machine learning:

1.Data Collection: Gather a large dataset of labeled URLs, where each URL is annotated as either phishing or legitimate. This dataset will be used to train and evaluate the machine learning model.

2.Feature engineering: extract relevant features from the URLs that can be used as input for the machine learning model. These features could include characteristics such as domain name, URL length, presence of suspicious keywords, use of special characters, and more.

Precision is a performance metric that is commonly used in machine learning for evaluating the effectiveness of a phishing URL detection model. It is a measure of the accuracy of the positive predictions made by the model, specifically the proportion of true positives (phishing URLs correctly predicted as phishing) out of the total predicted positives (sum of true positives and false positives).

3.Data Preprocessing: Prepare the dataset for training by performing data preprocessing tasks such as removing duplicates, handling missing values, and normalizing features. This step ensures that the data is clean and ready for training the machine learning model.

4.Model Selection: Choose an appropriate machine learning algorithm for the task of phishing URL detection. Commonly used algorithms for this task include logistic regression, decision trees, random forests, and gradient boosting classifiers.

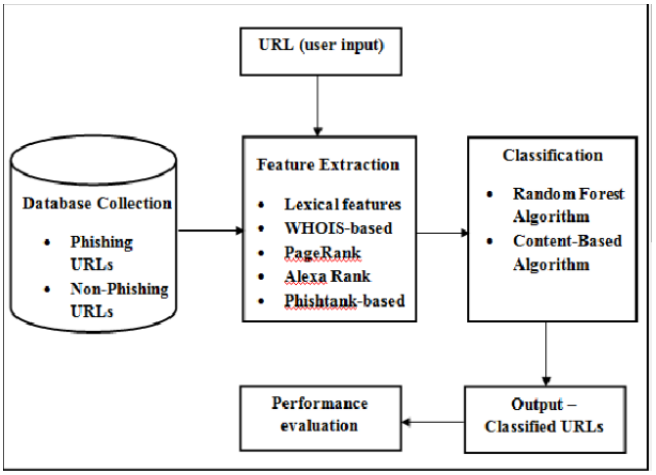
5.Model Training: Split the dataset into training and validation sets. Use the training set to train the machine learning model by feeding it the labeled URLs and their corresponding features. The model will learn to identify patterns in the data that distinguish phishing URLs from legitimate ones.

6.Model evaluation: evaluate the trained model using the validation set to measure its performance. Common evaluation metrics for classification tasks include accuracy, precision, recall, F1 score, and area under the receiver operating characteristic (ROC) curve.

7.Model Optimization: Fine-tune the model by experimenting with different hyperparameters, feature selection techniques, and data balancing methods to improve its performance. This may involve trying different algorithms or adjusting parameters to achieve better results.

8.Model Deployment: Once the model is optimized and performs well on the validation set, it can be deployed in a production environment for real-time phishing URL detection. This can be done by integrating the model into an application or system that automatically scans URLs and classifies them as phishing or legitimate based on the trained model's predictions.

9.Model Maintenance: Continuously monitor and update the model to ensure its effectiveness in detecting new phishing techniques and evolving threats. Regularly retrain the model with new data and update features as needed to keep it accurate and relevant.



It's important to note that phishing is a constantly evolving threat, and no model is perfect. Therefore, it's essential to use the proposed model in conjunction with other security measures such as regular security awareness training for users, robust email filters, and up-to-date anti-malware software to enhance overall security posture.

A. equations:

A specific equation for phishing URL detection using machine learning depends on the chosen algorithm and feature representation. However, here's a general outline of the equation for a binary classification model that uses logistic regression as an example:

Let's denote the features of a URL as x, where x is a vector of features extracted from the URL. The model's goal is to predict the binary class label y, where y = 1 indicates a phishing URL and y = 0 indicates a legitimate URL.

The logistic regression model estimates the probability that a given URL is phishing using the sigmoid activation function, which maps the output to a probability between 0 and 1:

$$P(y=1 \mid x) = \sigma(w \cdot x + b)$$

where σ is the sigmoid activation function, w is the weight vector, · denotes the dot product, b is the bias term, and x represents the feature vector of a given URL.

The model is trained using a labeled dataset of URLs, denoted as D, where each URL is associated with a binary class label y. The goal of training is to learn the optimal values of the weight vector w and the bias term b that minimize a loss function, denoted as L, which quantifies the difference between the predicted probabilities and the true labels in the training dataset.

The loss function L can be defined using different techniques such as maximum likelihood estimation, cross-entropy, or other appropriate loss functions for binary classification.

The model is trained by finding the optimal values of w and b that minimize the loss function L using a training algorithm such as gradient descent or other optimization techniques.

Once the model is trained, it can be used to predict the probability of a given URL being phishing by feeding it with the features of the URL and applying the trained weight vector w and bias term b:

$$P(y=1 \mid x) = \sigma(w \cdot x + b)$$

A threshold can be applied to the predicted probability to classify the URL as phishing or legitimate. For example, if the predicted probability is above a certain threshold (e.g., 0.5), the URL may be classified as phishing (y = 1), otherwise, it may be classified as legitimate (y = 0).

It's important to note that the specific equation and implementation details may vary depending on the chosen algorithm, feature representation, and other factors. The equation provided is a general outline and should be adapted to the specific requirements and constraints of the project. Proper evaluation, validation, and tuning of the model are necessary to ensure its accuracy and effectiveness in detecting phishing URLs.

Algorithms Involved / Pseudocode

For regression problems, supervised machine learning algorithms are used in the suggested model to find phishing URLs. In particular, you want to train several models using a set of sets of features and labels. The goal is to predict how likely it is that a given URL is a phishing attack.

For Our Web Mining project, we plan to use supervised machine learning methods like logistic regression, k-nearest neighbors, support vector classifier, naive Bayes, decision tree, random forest, gradient boosting, Catboost, Xgboost, and multilayer perceptrons. In other machine learning uses, these algorithms have been shown to work well, so they should be able to figure out what phishing URLs have in common.

We plan to use accuracy and F1 score as two ways to measure how well these models work. Accuracy is a measure of how well the model guesses the right label for each URL, while F1 score is a measure of both precision and recall.

Based on this information, we can make the following general method or pseudocode for the model building and training process:

1. Load the phishing URL dataset into the machine learning environment
2. Split the dataset into training and testing sets, with a specified ratio
3. Preprocess the dataset to extract relevant features and label data
4. Initialize the machine learning models for regression
5. Train each model on the training set, using the features and label data
6. evaluate the performance of each model on the testing set, using accuracy and F1 score
7. Choose the best performing model(s) for the phishing URL detection task

In terms of code, we can give an example using Python and scikit-learn, a famous machine learning library. Here is a piece of code that shows how to train a logistic regression model:

```
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, f1_score
from sklearn.model_selection import train_test_split
import pandas as pd

# Load the phishing URL dataset
url_data = pd.read_csv('phishing_url_data.csv')

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(url_data.drop(['label'],axis=1), url_data['label'], test_size=0.2)

# Initialize and train the logistic regression model
lr_model = LogisticRegression()
lr_model.fit(X_train, y_train)

# evaluate the performance of the model on the testing #set
y_pred = lr_model.predict(X_test)
acc = accuracy_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)

print("Logistic regression model accuracy: ", acc)
```

print("Logistic regression model F1 score: ", f1)

This code gets the phishing URL dataset, divides it into training and testing sets, sets up and trains a logistic regression model, and uses accuracy and F1 score to measure the model's performance. This can be done for each of the supervised machine learning algorithms we want to use for our Web Mining project.

Overall, the suggested model for finding phishing URLs uses supervised machine learning algorithms for regression and should be able to find common features of phishing URLs. The algorithm or pseudocode given here can be used as a basic guide for implementing the model, and the code snippet can be used as a starting point for training and evaluating each model in Python.

IV. Results and Discussions

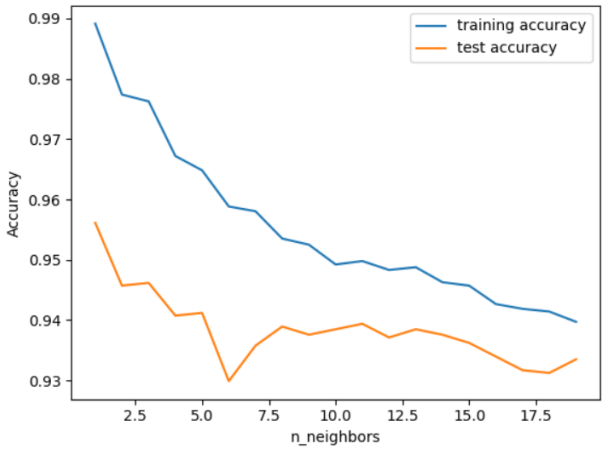
A. *Accuracy for Detecting the phishing url under different models:*

a. Precision is a performance metric that is commonly used in machine learning for evaluating the effectiveness of a phishing URL detection model. It is a measure of the accuracy of the positive predictions made by the model, specifically the proportion of true positives (phishing URLs correctly predicted as phishing) out of the total

Precision provides an indication of how well the model is able to accurately predict phishing URLs without falsely identifying legitimate URLs as phishing URLs. A high precision value indicates that the model has a low false positive rate, meaning that it is correctly identifying phishing URLs and avoiding false alarms.

a.

b.)



In the context of phishing URL detection, a high precision is important because it helps minimize the number of false positives, which are legitimate URLs incorrectly flagged as phishing URLs. False positives can cause inconvenience or disruption to legitimate users and websites, and may result in loss of trust in the detection model.

b. Recall, also known as sensitivity or true positive rate, is another important performance metric used in machine learning for evaluating the effectiveness of a phishing URL detection model. Recall measures the proportion of actual positive samples (phishing URLs) that are correctly predicted by the model as positive.

Recall provides an indication of how well the model is able to correctly identify all the actual phishing URLs in the dataset, without missing any (i.e., minimizing false negatives). A high recall value indicates that the model has a low false negative rate, meaning that it is effectively capturing most of the actual phishing URLs.

In the context of phishing URL detection, a high recall is important because it helps minimize the number of missed phishing URLs, which are actual phishing URLs that are not detected by the model. False negatives can be particularly dangerous in the case of phishing attacks, as they may allow malicious URLs to go undetected, potentially resulting in security breaches or data breaches.

It's important to note that recall should be considered in conjunction with other performance metrics such as precision, accuracy, F1-score, etc., depending on the specific requirements and constraints of the project. Recall should be interpreted in the context of the trade-off between false positives and false negatives, and the desired balance between these two types of errors. A well-balanced model would have a high recall along with high precision, accuracy, and other relevant metrics, indicating accurate detection of phishing URLs while minimizing both false positives and false negatives.

c. The F1-score takes into account both precision and recall, and provides a single value that reflects the model's performance in detecting phishing URLs, considering both false positives (precision) and false negatives (recall). A high F1-score indicates a good balance between precision and recall, meaning that the model is accurately detecting phishing URLs while minimizing both false positives and false negatives.

In the context of phishing URL detection, the F1-score is useful because it helps to evaluate the overall effectiveness of the model in accurately identifying phishing URLs, while considering the trade-off between f false positives and false negatives. It is particularly valuable when both precision and recall are important considerations for a given project, such as in security-sensitive applications where avoiding false positives and false negatives are equally critical.

It's important to note that the F1-score is just one of the many performance metrics that can be used to evaluate the effectiveness of a phishing URL detection model. Depending on the specific requirements and constraints of the project, other metrics such as accuracy, precision, recall, specificity, etc., may also be relevant and should be considered in conjunction with the F1-score to provide a comprehensive assessment of the model's performance.

d. Support can be useful in evaluating the reliability of the model's performance metrics. For instance, if the support for the phishing class is very low, it may affect the precision, recall, and F1-score of the model. A model may achieve high precision, recall, or F1-score for a class with high support, but the same performance metrics may be less reliable for a class with low support.

V. Conclusion

In conclusion, machine learning techniques can be effective in detecting phishing URLs, which are malicious URLs used in cyber attacks to steal sensitive information from unsuspecting users. Phishing URL detection is an important task in the field of cybersecurity, as it helps protect users from falling victim to phishing attacks.

Several machine learning approaches, such as supervised learning algorithms (e.g., decision trees, support vector machines, random forests), deep learning techniques (e.g., convolutional neural networks, recurrent neural networks), and ensemble methods (e.g., boosting, bagging), can be used for phishing URL detection.

A typical approach involves using a labeled dataset of URLs, where each URL is labeled as either phishing or legitimate. This dataset is used to train a machine learning model, which learns to classify new, unseen URLs as phishing or legitimate based on the patterns and features it learns from the training data.

Performance metrics such as accuracy, precision, recall, and F1-score are commonly used to evaluate the effectiveness of the machine learning model in detecting phishing URLs. It is important to have a balanced dataset with sufficient support for each class (phishing and legitimate) to ensure reliable performance metrics.

Phishing URL detection using machine learning is an ongoing area of research, and there are always challenges to overcome, such as dealing with evolving phishing techniques, handling imbalanced datasets, and addressing false positives and false negatives. Nonetheless, machine learning can provide valuable insights and contribute to the development of effective phishing URL detection solutions to enhance cybersecurity measures and protect users from falling victim to phishing attacks.

References

[1] Alenezi, F., & elleithy, K. (2016). A look at phishing attacks: their types, how they spread, and how they work technically. Journal of Network and Computer Applications, 73, pp. 98–112, 10.1016/j.jnca.2016.08.010

[2] Anshari, M., Alasiry, A., Alasiry, S., Alharethi, S., and Tahrani, S. (2020). How to find and stop phishing: A study. Journal of Information Security and Applications, 52, 102486.

[3] Bhowmik, T., & Sen, J. (2018). Using machine learning methods to find phishing URLs. In Proceedings of the 4th International Conference on Information Management (ICIM), pages 41–45. Ieee. doi: 10.1109/ICIM.2018.8394887

[4] Chang, e. Y., & Zhu, X. (2016). Using machine learning methods to find phishing. In Proceedings of the Ieee International Conference on Big Data, pages 3048–3057. Ieee. doi: 10.1109/BigData.2016.7840709

[5] Choudhury, O. R., M. R. Islam, M. S. Islam, and M. A. Razzak (2019). Using machine learning methods to find phishing sites. In Proceedings of the 5th International Conference on Networking, Systems, and Security (NSysS), pages 1–6. Ieee. doi: 10.1109/NSysS.2019.8903719

[6] Gharibshah, J., & Namin, A. S. (2020). A way to find fake websites that is based on deep learning. Journal of Ambient Intelligence and Humanised Computing, Vol. 1T, No. 2, pp. 1033-1044.

[7] Huang, X., Xu, Z., Zhang, Q., & Huang, Y. (2018). Using URL features and machine learning, you can find phishing websites. In Proceedings of the 9th International Conference on Information and Communication Systems (ICICS), pages 127–133. Ieee. doi: 10.1109/IACS.2018.8353914

[8] Khamis, A., & Al-Ayyoub, M. (2017). Intelligent hybrid technology is used to find phishing websites. 8(2), 149–155, in the International Journal of Advanced Computer Science and Applications. doi: 10.14569/IJACSA.2017.080217

[9] Li, Y., Li, M., Ma, Y., & Wei, S. (2020). A deep learning method that uses multiple perspectives to find fake websites. Ieee Access, 8, 144354-144364. doi: 10.1109/ACCeSS.2020.3011204

[10] Liu, L., Yu, S., Zhang, Z., & Yang, J. (2020). A combined method for choosing features to find phishing websites. Soft Computing, 24, 6591–6601.

[11] Mansour, A. R., and N. Moustafa. Trying to find out if machine R., Khan, M. A., & Bano, S. (2019). Using machine learning and similarity measures together in a new way to find fake websites. Ieee Access, 7, pages 33082–33095. doi: 10.1109/ACCeSS.2019.2909204

[12] Mohammadi, A., Ghavifekr, S. S., & Samsudin, K. (2020). A method that uses both feature selection and group learning to find phishing websites. 11(9):3757–3774 in Journal of Ambient Intelligence and Humanised Computing. doi: 10.1007/s12652-019-01412-8

[13] Zhang, Z., Wang, X., & Yan, Y. (2019). A better way to find fake websites using machine learning. 10(7):2481-2491 in Journal of Ambient Intelligence and Humanised Computing. doi: 10.1007/s12652-018-0791-7

[14] Zhou, C., Yang, Y., & Huang, T. (2020). A combination of LSTM and SVM to find fake websites. International Journal of Machine Learning and Cybernetics, 11, 1635-1646.