Detecting Phishing Websites Using Machine Learning Techniques

CENG 3544, Computer and Network Security

Selman Olgun
selmanolgun@posta.mu.edu.tr

Monday 24th June, 2024

Abstract

This study aims to detect phishing websites using machine learning techniques. The dataset is trained with various machine learning models and their performances are compared. Logistic Regression, KNN, Random Forest, Naive Bayes and Decision Tree models were used and the Stacking model, which is an Ensemble method, was created by selecting the models (Random Forest, Logistic Regression, Decision Tree) with the best results. Using this model, we predict whether the URLs entered by users are phishing or not. The results show that the Stacking model provides high accuracy and reliability.

1 Introduction

Phishing attacks are a method of creating malicious sites that mimic legitimate websites in order to steal users' personal and financial information. Phishing is a fraud technique that aims to steal users' identity and financial information using social and technological tricks, using a variety of methods such as email, URLs, messaging and forum posts [1]. Such attacks pose a major threat to internet users and lead to serious information loss.

Machine learning has proven effective in detecting phishing websites by extracting significant patterns from extensive datasets [2]. In this study, we aim to detect phishing websites using machine learning techniques.

In our study, the "Phishing Websites" dataset from the UCI Machine Learning Repository was used. The web_traffic and page_ranking columns in the dataset were deleted because they contained outdated information. Logistic Regression, KNN, Random Forest, Naive Bayes and Decision Tree models were trained using the remaining features. The performance of the models was evaluated by F1 score and accuracy metrics, and the three best performing models (Random Forest, Logistic Regression, Decision Tree) were selected and an Ensemble method Stacking model was created.

This study makes an important contribution to the detection of phishing websites with machine learning techniques in order to increase the security of internet users.

2 Fundamentals

In this section, the basic concepts and machine learning algorithms used in the study are explained. Information about the features and machine learning techniques used for the detection

of phishing websites will be provided.

2.1 Phishing

Phishing is a method of creating malicious websites that mimic legitimate websites in order to steal users' personal information. These attacks often send fake links via email or text message, misleading users and getting them to share sensitive information. Phishing attacks are social engineering crimes where users receive fraudulent emails asking them to provide their credentials to a trusted organization [3]. However, in this scenario, a deceitful entity pretends to be a reputable organization with the goal of obtaining users' financial or personal information unlawfully. [4]. Phishing poses a serious threat to internet users and can cause financial losses.

2.2 Machine Learning

Machine learning falls under artificial intelligence, enabling computers to learn from data and utilize that learning to forecast results for new data. Machine learning algorithms allow new data to be classified by recognizing patterns in the data and using these patterns. The use of machine learning algorithms, especially for phishing domain detection, stands out as an effective method in solving categorization problems. Machine learning algorithms, especially for phishing domain detection, stand out as an effective method in solving categorization problems [5].

2.3 Feature Extraction

Feature extraction is the process of extracting meaningful information from raw data. In this study, various features were extracted from URLs. The phishing websites have certain characteristics and patterns that can be considered as features [6]. For instance, factors such as URL length, the inclusion of special characters, and the presence of an IP address are utilized in identifying phishing websites. These features play an important role in training machine learning models.

2.4 Stacking

Stacking is an Ensemble learning technique and is used to make a final prediction by combining the predictions of different machine learning models. Unlike bagging and boosting algorithms, which aggregate individual learners of the same type, the stacking algorithm typically combines single learners constructed by different learning algorithms.[7] [8]. In this study, a Stacking model is created using the three best performing models (Random Forest, Logistic Regression, Decision Tree).

3 Related Works

The dataset and feature extraction rules used in this study are inspired by Rami Mohammad and Lee McCluskey's "An Assessment of Features Related to Phishing Websites using an Automated Technique". Mohammad and McCluskey's work focuses on the automatic extraction of important features for the detection of phishing websites. In their work, several features that are effective in identifying phishing sites are identified and tested experimentally. In particular, features such as URL length, the presence of IP address, the use of the '@' symbol in the URL formed the basis of this work [9].

Mohammad and McCluskey's work is not only an important reference for phishing detection, but also a guide for automatic feature extraction. The rules used in their work are similarly applied in our project and integrated with machine learning models. Furthermore, research published in IEEE Xplore demonstrates the use of comparable machine learning algorithms for detecting phishing attacks. [10]. Unlike this work, we use various machine learning models to select the best performing models and combine them together to form the Stacking model, which is an Ensemble method.

4 System Proposal and Implementation

In this section, we detail the proposed solution for detecting phishing websites and how it is implemented. We adapted various appliance intelligence models resorting to the dataset. The models used are: data preparation and feature extraction, and training and evaluation of machine learning models.

4.1 Data Preparation and Feature Extraction

The dataset used in the proposed solution is the "Phishing Websites" dataset from the UCI Machine Learning Repository. This dataset was created by extracting various features from URLs. The web_traffic and page_ranking columns were removed because they contain outdated information. The remaining features are used to determine whether URLs are phishing or not.

During the feature origin process, differing countenance in the way that the time of the URLs, thevicinity of an IP address, the use of the '@' letter in the URL were without thinking elicited. These features provide important inputs for training machine learning models. The rules used for feature extraction are in line with the rules outlined in Mohammad and McCluskey's work [9]. Furthermore, the success of the model is highly dependent on the quality of the datasets used.

The dataset is detached into preparation and test sets and the dossier is scaled.. The scaling of the features was done to improve the performance of the models and to facilitate training.

4.2 Training Machine Learning Models

We prepared various machine intelligence models utilizing the dataset. These models include:

- Logistic Regression
- K-Nearest Neighbors (KNN)
- Random Forest
- Naive Bayes
- Decision Tree

The performance of the models was evaluated using F1 score and accuracy metrics. The three best performing models (Random Forest, Logistic Regression, Decision Tree) were selected and We integrated these models to create the Stacking model, which is an ensemble method..

4.3 Stacking Model

Stacking is an Ensemble learning technique used to make a final prediction by combining the predictions of different machine learning models. In this method, the predictions of the first level models are combined and the final prediction is made. In this study, a Stacking model was created using the three best performing models (Random Forest, Logistic Regression, Decision Tree). The stacking model is trained with training data and evaluated on test data. The performance of the model is evaluated by the F1 score and accuracy rate calculated on the test set.

5 Results and Discussion

The experimental findings indicate that the Stacking model achieves superior accuracy and F1 score. This proves how effective machine learning models are in detecting phishing websites.

5.1 Performance Evaluation of Models

The performance evaluation of the applied machine learning models was performed using F1 score and accuracy metrics. Table 1 shows the performance of each model. The best performing model was the Stacking model.

Table 1: Performance of Machine Learning Models

Model	F1 Score	Accuracy
Stacking	0.964	0.959
Random Forest	0.962	0.956
Decision Tree	0.957	0.951
Logistic Regression	0.939	0.930
KNN	0.938	0.929
Naive Bayes	0.399	0.573

5.2 Analysis of Results

The experimental results show that the Stacking model provides higher accuracy and F1 score than the other individual models. This proves that combining the strengths of different models is effective in improving the overall performance.

Random Forest and Decision Tree models also showed high performance, but the Stacking model achieved better results by combining the predictions of these models. Logistic Regression and KNN models also performed acceptably, but the Naive Bayes model performed poorly.

These results show that machine learning models are effective in detecting phishing websites and the Stacking model provides higher accuracy in this area. The success of the Stacking model is due to the accuracy of feature extraction and the power of the selected models.

6 Conclusion

This study aims to discover phishing websites utilizing machine learning methods. Various machine learning models were trained on the "Phishing Websites" dataset from the UCI Machine

Learning Repository and the best performing models (Random Forest, Logistic Regression, Decision Tree) were selected and a Stacking model was created.

The exploratory results show that the Stacking model specifies bigger veracity and F1 score than the other individual models. This proves that combining the predictions of different models is effective in improving the overall performance.

The main contributions of the study are:

- Features used in the detection of phishing websites have been identified through automatic feature extraction.
- Various machine learning models are used for performance evaluation and the best models are selected.
- A stacking model is constructed by combining the best performing models and the superior performance of this model is demonstrated.

Future work could include adding new features to further improve the performance of the model and experimenting on larger datasets. Also, developing real-time phishing detection systems and deploying these models more widely could be an important step for future work.

In conclusion, this study has shown how effective machine learning techniques are in detecting phishing websites and has made significant contributions in this field.

References

[1] Dutta, A. K. (2021). Detecting phishing websites using machine learning technique. *PloS One, 16*(10), e0258361. Public Library of Science.

[2] Burbela, K. (2023). Model of detection of phishing URLsbased on machine learning. *DiVA Portal*.

[3] Gupta, B. B., Arachchilage, N. A., & Psannis, K. E. (2018). Defending against phishing attacks: taxonomy of methods, current issues and future directions. *Telecommunication Systems, 67*, 247-267.

[4] Alluwaici, M., Junoh, A. K., AlZoubi, W. A., Alazaidah, R., & Al-luwaici, W. (2020). New features selection method for multi-label classification based on the positive dependencies among labels. *Solid State Technology, 63*(2s).

[5] Basnet, R. B., Sung, A. H., & Liu, Q. (2011). Rule-based phishing attack detection. In *International conference on security and management (SAM 2011), Las Vegas, NV*.

[6] Alswailem, A., Alabdullah, B., Alrumayh, N., & Alsedrani, A. (2019, May). Detecting phishing websites using machine learning. In 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS) (pp. 1-6). IEEE.

[7] Lateko, A. A., Yang, H. T., Huang, C. M., Aprillia, H., Hsu, C. Y., Zhong, J. L., & Phng, N. H. (2021). Stacking ensemble method with the RNN meta-learner for short-term PV power forecasting. Energies, 14(16), 4733.

[8] Li, Y., Yang, Z., Chen, X., Yuan, H., & Liu, W. (2019). A stacking model using URL and HTML features for phishing webpage detection. *Future Generation Computer Systems*.

[9] Mohammad, R. M., & McCluskey, L. (2012). An Assessment of Features Related to Phishing Websites using an Automated Technique. In *The 7th International Conference for Internet Technology and Secured Transactions (ICITST-2012)*.

[10] Alrefaai, S., Özdemir, G., & Mohamed, A. (2022, June). Detecting phishing websites using machine learning. In 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) (pp. 1-6). IEEE.