NFSU JOURNAL OF CYBER SECURITY & DIGITAL FORENSICS

NFSU – Journal of Cyber Security and Digital Forensics
Volume – 2, Issue – 2, December 2023
E – ISSN – 2583-7559

EXPLORING FEATURE IMPORTANCE IN PHISHING URL DETECTION MODELS

Shaurya

School of Cyber Security and Digital Forensic,
National Forensic Sciences University,
Gandhinagar, India
shaurya1337@gmail.com

Dr. Ravirajsinh S. Vaghela

School of Cyber Security and Digital forensics,
National Forensic Sciences University,
Gandhinagar, India
ravirajsinh.vaghela@nfsu.ac.in

***Abstract*— Cybersecurity faces persistent threats from phishing attacks, prompting the need for robust URL detection systems. This research explores the efficacy and interpretability of Random Forest and Artificial Neural Network (ANN) models, employing SHAP (SHapley Additive exPlanations) for feature importance analysis. Using these models, phishing URLs were classified, and SHAP facilitated the understanding of feature significance in model decision-making. Comparative analysis revealed distinct feature preferences between models. Random Forest emphasized Google index, page rank, and web traffic, while ANN prioritized page rank, Google index, and URL structure attributes. These findings underscore the models' divergent feature inclinations, providing actionable insights for feature selection and model enhancement in phishing URL detection.**

Keywords*— *Phishing Detection, Machine Learning, Artificial Neural Networks, SHAP Analysis, Cybersecurity

INTRODUCTION

Cyber security stands as an ever-evolving battlefield where adversaries continuously craft sophisticated methods to exploit vulnerabilities, with phishing attacks representing a pervasive threat. Phishing, often executed through deceptive URLs, poses significant risks to individuals, organizations, and the integrity of digital infrastructures. Detecting these malicious URLs remains a critical challenge in safeguarding against cyber threats.

Phishing poses a significant security risk by employing advanced psychological and social manipulation methods to deceive people into clicking on harmful website links and sharing extremely valuable sensitive data, including personal or business-related details and account login credentials [1]. The term 'phishing' cleverly echoes 'fishing,' illustrating the deceptive tactics used to entice unaware individuals into disclosing sensitive information or engaging in actions that jeopardize their safety.

The roots of phishing can be traced back to the early 1990s when the internet was gaining prominence in both personal and professional spheres. One of the earliest documented instances dates to 1995, where attackers impersonated America Online (AOL) representatives, prompting users to disclose their login credentials via instant messages—a precursor to modern phishing techniques. The early 2000s witnessed a surge in phishing attacks facilitated by the proliferation of email services and the adoption of online banking and e-commerce.

According to the India Cyber Threat Report 2023 by the Data Security Council of India (DSCI), phishing remains a significant threat [2]. For the second year in a row, phishing was the leading infection vector, identified in 41% of incidents. More than half of phishing attacks used spear phishing attachments [3]. Phishing methods diversified beyond email-based schemes. Smishing (phishing via SMS or text messages), vishing (voice phishing via phone calls), and spear phishing (targeted attacks on specific individuals or organizations) emerged as variants, each leveraging different communication channels and tactics to achieve nefarious objectives.

Artificial Intelligence (AI) plays a pivotal role in modernizing and fortifying phishing detection methods due to its ability to process vast amounts of data, recognize intricate patterns, and adapt to evolving threats.

Some key aspects highlighting the importance of AI in phishing detection are as follows:

1. Adaptive and Dynamic Detection: AI-powered systems excel in adapting to the changing landscape of phishing attacks. Machine learning algorithms can learn from new data and patterns, continually evolving to detect novel phishing techniques that traditional rule-based systems might miss.
2. Pattern Recognition and Anomaly Detection: AI models, such as neural networks and ensemble methods like Random Forests, excel in pattern recognition. They can identify subtle anomalies within URLs, emails, or user behavior that might indicate a phishing attempt, even when the attack methods evolve or disguise themselves.
3. Real-Time Analysis: AI algorithms can swiftly analyze vast amounts of data in real-time. This capability is crucial in swiftly flagging suspicious URLs or emails, mitigating potential threats before they cause harm.
4. Automation and Scalability: AI-powered phishing detection systems automate the process of analyzing and flagging potential threats, freeing up human resources and enabling scalability to handle the increasing volume and complexity of attacks.

5. Interpretability and Explainability: Explainable AI techniques, like SHAP (SHapley Additive exPlanations), help in understanding and interpreting the decisions made by AI models. This aids in building trust and understanding how the system identifies phishing attempts, enhancing transparency and allowing for better refinement.

Explainable AI (XAI) serves as a critical tool in unraveling the complexities of feature weightage within artificial intelligence models, in domains like phishing detection. In the realm of cyber security, where the stakes are high and model interpretability is paramount, understanding feature weightage through XAI promotes better understanding and acceptance of these AI-driven systems. Stakeholders, including security analysts and end-users, gain insights into why a particular URL is flagged as malicious or benign, enhancing their ability to validate and trust the model's outcomes. XAI techniques, such as SHAP (SHapley Additive exPlanations), provide a structured framework to dissect and interpret the contributions of individual features, enhancing the transparency and interpretability of these models. Understanding AI model decisions and creating explanations that laypeople can grasp is crucial, making Explainable Artificial Intelligence (XAI) methods essential [4].

Machine learning models, particularly Random Forest and Artificial Neural Network (ANN) models using Keras, have shown promise in their ability to discern phishing URLs. However, comprehending the decision-making process of these models and understanding the significance of individual features in distinguishing between legitimate and malicious URLs are paramount for building robust and interpretable cybersecurity solutions.

This research endeavors to address this imperative need by employing SHAP (SHapley Additive exPlanations), an explainable AI algorithm, to unravel the feature importance within Random Forest and ANN Keras models for phishing URL detection. The primary objective is to compare and contrast the significance of features in these models, aiming to identify the key discriminative factors driving their decision-making process. This paper also deep diving into the identification of Discriminative features. Discerning the most influential features that contribute to the models' predictive capabilities in distinguishing phishing URLs. Highlighting the practical implications of understanding feature importance in prioritizing feature extraction and enhancing model interpretability in real-life cyber security applications.

RELATED WORK

According to M. Vijayalakshmi [5] Phishing detection techniques encompass a range of strategies employed to identify and mitigate phishing threats. These techniques can be extensively classified into three categories:

List-based techniques rely on maintaining databases or lists of known phishing URLs or associated characteristics. Local whitelists and blacklists have been used in numerous research studies to prevent falling victim to phishing scenarios. While the list-based approach is capable of finding the malicious URLs faster than other approaches, its detection rate is not as good as other approaches. This is a result of the blacklists not being updated regularly [6]. Heuristic rule-based techniques involve creating rules or heuristics based on observable patterns or characteristics commonly found in phishing URLs. To detect web phishing attacks based on URLs, Sahingoz et al. [7] employed heuristics to extract natural language processing (NLP) attributes from the URLs. These heuristics were designed considering elements such as raw word count, short word length, Alexa rating, occurrences of similar brand names, among other factors. Learning-based techniques leverage machine learning algorithms to detect phishing attempts by analyzing features extracted from URLs, email content, or user behavior. In order to identify web phishing attacks and extract naturally occurring properties from URLs, Yang et al. [8] suggested a deep learning approach. This approach employs a long short-term memory (LSTM) network to understand the sequential relationships within character sequences and utilizes a convolutional neural network (CNN) to identify and extract correlated features.

The writers of [9] examined a number of phishing webpage characteristics and identified the top 19 qualities. Following the extraction of these 19 variables from the webpage's source code, they classified phishing websites with almost 99% accuracy using support vector machines (SVM), random forests (RF), neural networks, logistic regression, and Naïve Bayes (NB).

An autonomous intelligent method for phishing webpage detection was presented by Xiaoqing et al. [10]. They employed NB for classification after analyzing the characteristics of the uniform resource location (URL). SVM is used to parse and reclassify websites that seem suspect. Based on their findings, they assert that the system provides excellent detection accuracy in a shorter amount of time.

A phishing detection method built by integrating the webpage's URL and source code was described by Wu et al. [11]. In their suggested approach for phishing webpage identification, they employed SVM as a machine learning model and the Levenshtein method for determining string similarity.

The Random Forest from Trees learning approach demonstrated the greatest results in terms of Accuracy and TP rate, with 97.259% and 0.973, respectively, according to a paper by R. Alazaidah [12].

Collectively, these studies underscore the efficacy of diverse AI algorithms, including SVM, RF, NB, and ensemble methods, in detecting phishing attempts. Their high accuracy rates and ability to process various characteristics, whether from URL attributes or webpage source codes, emphasize the potential for multi-faceted approaches in constructing robust and efficient phishing detection systems.

NFSU JOURNAL OF CYBER SECURITY & DIGITAL FORENSICS

NFSU – Journal of Cyber Security and Digital Forensics
Volume – 2, Issue – 2, December 2023
E – ISSN – 2583-7559

SHAP VALUES IN MODEL INTERPRETATION

Shapley values, rooted in game theory, have gained prominence for their role in fair credit allocation among participants in a cooperative game. Specifically, in the context of Explainable AI (XAI), SHAP (SHapley Additive exPlanations) emerges as a powerful method proposed by Lundberg and Lee [13] for explaining individual predictions made by machine learning models. SHAP, short for SHapley Additive exPlanations, stands out as a widely recognized visualization tool within explainable artificial intelligence algorithms. It assists in offering comprehensive explanations of prediction models, allowing examination of each predictor's contribution to the final output. [14].

In explainable artificial intelligence (XAI), a machine learning model that operates as an opaque system, with complex and inaccessible internal mechanisms, is termed a "black-box model." These models utilize input information to make predictions, yet users are unable to access the decision-making process or the rationale behind these predictions [15]. The lack of transparency in "black-box" AI systems and certain algorithms presents substantial ethical challenges, particularly concerning trust-related queries [16].

Black box models often lack transparency in how they arrive at their decisions. In cyber security, understanding why a model flagged a particular URL or email as malicious is crucial for validation and trust. SHAP values are crucial as they address the "black box" problem in complex AI models by revealing the contribution of each feature to the model's predictions, enhancing interpretability. These models often produce accurate predictions but lack transparency in revealing how they arrived at those decisions. SHAP values offer a solution by assigning each feature a numerical value, representing its contribution to each prediction. This allows practitioners and stakeholders to grasp not only which features were influential but also the direction and magnitude of their impact on the model's output.

SHAP values provide a detailed grasp of feature significance by measuring the influence of individual features on a model's predictions. In phishing detection, this means identifying which characteristics or attributes of URLs, such as domain structure, lexical cues, or content-related indicators, carry more weight in differentiating between legitimate and malicious instances.

The section encompasses three primary components: Dataset Description, Model Selection and Justification, and SHAP Computation. The Dataset Description provides an overview of the dataset used in this study, detailing its source, size, characteristics, and any preprocessing steps employed. Following this, the section on Model Selection and Justification outlines the rationale behind the selection of specific machine learning models for phishing detection, emphasizing their strengths and relevance in this context. Finally, the SHAP Computation section details the methodology and implementation of SHAP (SHapley Additive exPlanations), elucidating how this technique was employed to analyze the feature importance and interpretability of the selected models in phishing URL detection.

A. Dataset Selection

Choosing a database with pre-extracted features offers several advantages when conducting research, particularly in fields like phishing detection. Firstly, utilizing a dataset with pre-extracted features saves considerable time and resources. Feature extraction from raw data, especially in complex domains like cyber security, can be laborious and technically challenging. By using a dataset where this process has been completed, researchers can focus directly on model development, analysis, and validation, expediting the research timeline significantly.

Moreover, pre-extracted feature datasets frequently include features crafted by domain experts or through robust methodologies. These features might encapsulate nuanced aspects or characteristics specific to the problem domain, which could be challenging to extract comprehensively without expertise. Leveraging such features can enhance the performance and effectiveness of the models.

Furthermore, with pre-extracted features we can shift our emphasis towards leveraging SHAP analysis to delve deeper into the models' decision-making processes, dedicating more time to SHAP analysis allowing for a thorough validation of the model's behavior.

In this study, we utilized the database curated by H. Abdelhakim,[17] comprising 11,430 URLs, each characterized by 87 distinct features. This dataset serves as a benchmark for machine learning-based phishing detection systems, offering a robust collection of features essential for model training and evaluation. The dataset is structured with features categorized into three classes: 56 features extracted from URL structure and syntax, 24 features derived from the content of corresponding web pages, and 7 features obtained by querying external services. Notably, this dataset is meticulously balanced, comprising an equal distribution of 50% phishing URLs and 50% legitimate URLs, ensuring parity in representation. It's essential to note that these datasets were constructed in May 2020, serving as a contemporary foundation for our analyses and model development.

B. Model Selection and Justification

Selecting the right model for phishing detection involves considering various factors such as performance, interpretability, scalability, and adaptability to evolving threats. Two common models used in this domain are Random Forest and Artificial Neural Networks (ANN) implemented with Keras. Here’s a breakdown of their selection and justification:

1) *Random Forest:*

A random forest comprises 'n' decision trees, each of which generates distinct outputs for the same input [18]. Random Forest is a versatile and powerful machine learning

algorithm commonly used in various domains, including cyber security for phishing detection. It falls under the category of ensemble learning, where multiple decision trees are built and combined to make predictions. Here are the features of Random Forest:

a) Ensemble of Decision Trees: The Random Forest technique forms an ensemble of decision trees, where each tree is built using a portion of the training data and a subset of the available features. These trees operate independently and collectively contribute to the final prediction.

b) Bagging and Randomization: It employs a technique called bagging (bootstrap aggregating) to build individual trees. Additionally, it introduces randomness by using a random subset of features at each split in the trees, which helps to mitigate overfitting and promotes robustness.

c) Voting for Predictions: When making predictions, each decision tree in the forest produces its outcome. In classification tasks like phishing detection, the final prediction is determined by aggregating the votes (e.g., using majority voting) from all individual trees.

d) Rationale for Choosing Random Forest:

The selection of Random Forest for our phishing detection task is underpinned by several crucial considerations, affirming its suitability and advantages in handling the complexities inherent in our dataset and the SHAP (SHapley Additive exPlanations) interpretability technique.

e) Handling Large Number of Features: Our dataset comprises an extensive set of 87 features, encompassing various aspects crucial for phishing detection. Random Forest is particularly adept at managing high-dimensional data. Its ensemble nature, coupled with the technique of randomly selecting a subset of features at each split in the trees, enables efficient handling of a vast number of attributes without succumbing to overfitting.

f) Compatibility with SHAP for Interpretability: Utilizing SHAP, a powerful method for explaining the output of machine learning models, is pivotal in unraveling the black box of Random Forest. SHAP reveals feature importance, aiding in understanding each feature's contribution to predictions; it harmonizes well with Random Forest's capability to compute importance, enhancing interpretability in the model's decision-making process.

g. Advantages of Using Random Forest over other Models : In this investigative study [19], a comparative analysis employed four distinct machine learning methodologies (Support Vector Machines [SVM], Artificial Neural Networks [ANN], Random Forest [RF], and Decision Trees [DT]) within an experimental framework to ascertain the most precise machine learning model for identifying deceptive online domains. The Random Forest model exhibited the supreme accuracy in detection, yielding a rate of 97%, trailed closely by DT with 96%, ANN at 95%, and SVM achieving 94%.

2) *Artificial Neural Networks:*

Artificial Neural Networks (ANNs) is a popular method for handling difficult problems in the real world. Potential applications range from path planning to account formulae. It represents a greatly streamlined form of the nervous system in humans. Neural-like computational units comprise an artificial neural network (ANN). The input, output, and hidden layers make up the three layers of an ANN model in general [20].Artificial Neural Networks (ANNs) implemented through Keras, a high-level neural network API, stand as a powerful tool in phishing detection owing to their capacity to discern complex patterns within data. It serves as a high-level interface for building, training, and deploying artificial neural networks (ANNs) with simplicity and flexibility. The application of ANNs in this domain involves a structured framework that harnesses the capabilities of deep learning to identify subtle yet critical features indicative of phishing attempts.

Here are the features of Artificial Neural Networks:

a) Activation Functions: Complex challenges are typically associated with high dimensional nonlinear data. ANNs should leverage the nonlinear activation functions (AFs) of their hidden layers in order to learn such problems efficiently [21]. Keras offers a variety of activation functions like ReLU, Sigmoid, or Tanh, enabling the introduction of non-linearities within the network. This nonlinearity allows the model to learn complex, non-linear relationships within the data, crucial for discerning subtle differences between legitimate and phishing URLs.

b) Customizable Layers: Keras allows the construction of custom network architectures, enabling the design of networks tailored to specific requirements in phishing detection. This flexibility allows for experimentation with various layer configurations, optimizing the model's performance.

c) Learning Complex Patterns: ANNs, being highly adaptable, can learn intricate patterns inherent in phishing URLs, even in the presence of noise or variations. This adaptability allows for generalizing learned patterns to detect unseen phishing attempts, contributing to robust detection capabilities.

d) Rationale for Choosing Artificial Neural Networks:

The decision to incorporate Artificial Neural Networks (ANN) in our phishing detection task is grounded in several critical considerations that highlight its suitability and advantages, particularly concerning the complexities inherent in our dataset and the utilization of the SHAP (SHapley Additive exPlanations) interpretability technique.

e) Handling Large Number of Features: Just like Random Forest, ANNs, including those implemented using Keras, are scalable and can handle a large number of features. They can process diverse and extensive datasets, making them well-suited for our phishing detection scenario where up to 87 features need to be considered.

f) Compatibility with SHAP for Interpretability: Artificial Neural Networks (ANN) implemented with Keras present a significant challenge in terms of interpretability due to their inherent complexity. However, compatibility with SHAP (SHapley Additive exPlanations) is feasible and valuable in enhancing their interpretability.

3) Limitation

In the pursuit of expanding the research horizons and addressing the comprehensive landscape of phishing URL

detection, the study encountered a notable obstacle: the challenge of integrating additional algorithms, particularly Support Vector Machines (SVM), into the investigative framework. Since other researchers [23] have achieved a 95.80% recognition rate of phishing urls using Support Vector Machine Models hence the algorithm could have been a good choice for the experiment. This hurdle emerged from the absence of official support for SVM within the SHAP (SHapley Additive exPlanations) framework at the time of experimental stage. Despite persistent efforts to circumvent this limitation by attempting to integrate SVM independently, the endeavors were met with formidable barriers, primarily in the form of exceptionally prolonged training times.

PERFORMANCE TEST RESULTS

- *Random Forest*
  - *Evoulation Metrices*

```
Accuracy: 0.9658792650918635
Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.97      0.97      1157
           1       0.97      0.96      0.97      1129

    accuracy                           0.97      2286
   macro avg       0.97      0.97      0.97      2286
weighted avg       0.97      0.97      0.97      2286
```

Figure 2 Accuracy, Precision, Recall and F1 Score of Random Forest Model

As we be can see 97% percentage accuracy can be achieved here in model testing.

- *Confusion Matrix*

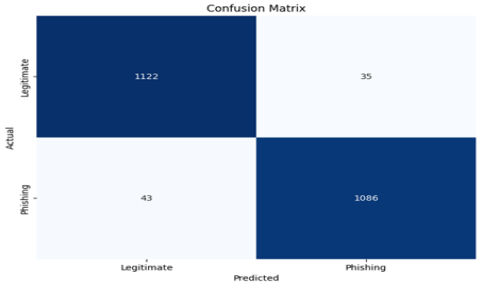

Figure 2 Confusion matrix

From figure 2. We can deduce Confusion matrix high True positive and false positive classified by Random forest model.

- *ROC Curve*

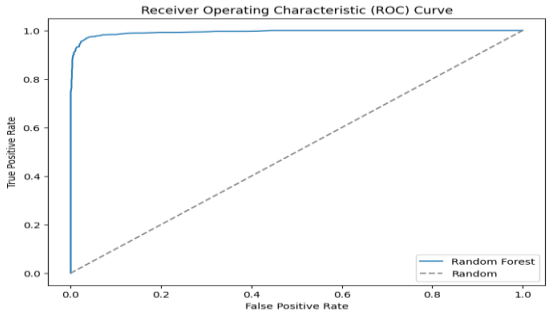

Figure 3 ROC Curve of Random Forest Model

Figure 3 ROC curve also give best result in Receiver operating charateristics True postiive rate good indicator.

- *SHAP Analysis of Random Forest Model*

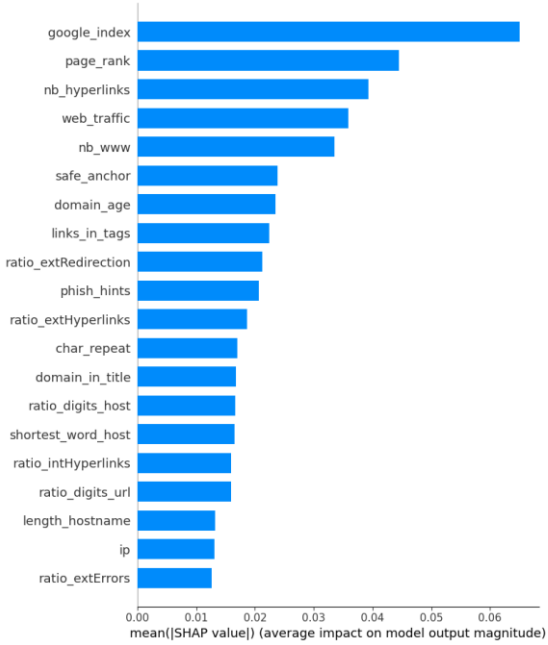

Figure 4. Bar Graph showcasing Top 20 Features

Based on SHAP analysis Figure 4. test it clearly indicate google_index and page_rank having highest SHAP value it means having which feature play crucial role in this predication.

- *SHAP Means Value*

Feature Names	Mean SHAP Value
google_index	0.06510
page_rank	0.04460
nb_hyperlinks	0.03940
web_traffic	0.03590
nb_www	0.03360
safe_anchor	0.02390
domain_age	0.02350
links_in_tags	0.02240
ratio_extRedirection	0.02130
phish_hints	0.02060
ratio_extHyperlinks	0.01870
char_repeat	0.01700
domain_in_title	0.01670
ratio_digits_host	0.01660
shortest_word_host	0.01650
ratio_digits_url	0.01590
ratio_intHyperlinks	0.01590
length_hostname	0.01320
ip	0.01310

Figure 5. Mean SHAP Values for top 20 Features

Based on SHAP mean analysis Figure 5. test it clearly indicate google_index and page_rank and nb_hyperlink having highest SHAP mean value it means having which feature play crucial role in this predication.

Mean absolute SHAP values were extracted for each feature and organized into a structured DataFrame, outlining the most influential features in predicting URL classifications. The top 20 most influential features are given in Fig.5 .

- *Artificial Neural Network*
  - *Evoulation Metrices*

```
Accuracy: 95.93%
              precision    recall  f1-score   support

           0       0.96      0.96      0.96      1157
           1       0.96      0.96      0.96      1129

    accuracy                           0.96      2286
   macro avg       0.96      0.96      0.96      2286
weighted avg       0.96      0.96      0.96      2286
```

Figure 6. Accuracy, Precision, Recall and F1 Score of ANN Model

As we be can see in Figure 6. 96% percentage accuracy can be achieved here in ANN model testing.

- *Confusion Matrix*

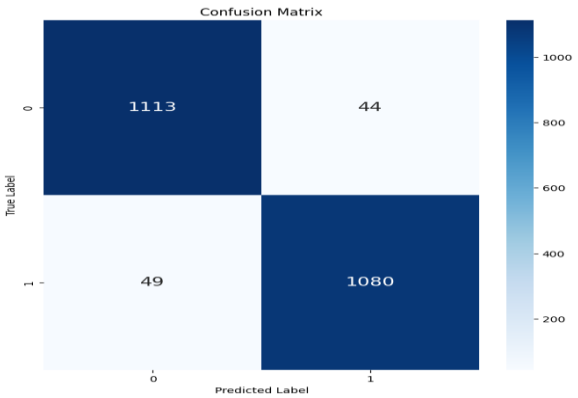

Figure 7. Confusion matrix

Figure 7. Confusion matrix high True positive and false positive classified by ANN model.

- *ROC Curve*

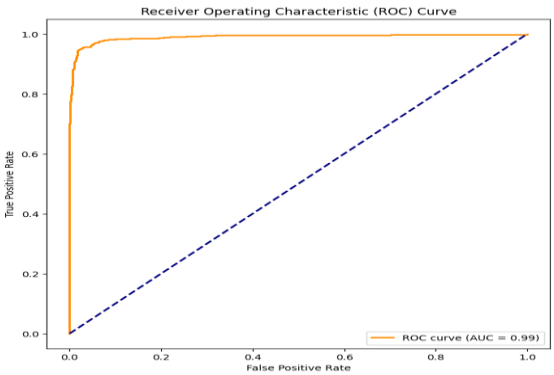

Figure 8 ROC Curve of ANN Model

Figure 8 ROC curve also give tuned result in Receiver operating charateristics True postiive rate good indicator for ANN model.

- *SHAP Analysis of Random Forest Model*

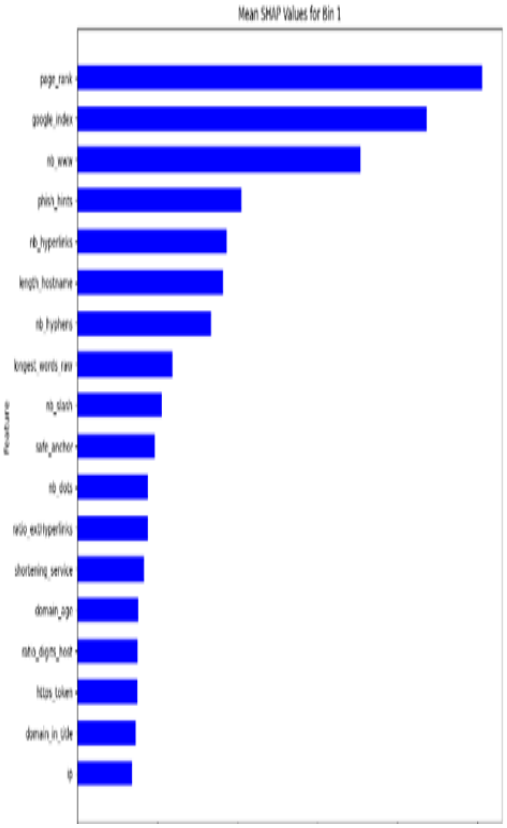

Figure 9. Bar Graph showcasing Top 20 Features

Based on SHAP analysis Figure 9. test it clearly indicate page rank and google_rank having highest SHAP value it means having which feature play crucial role in this predication

- *SHAP Means Value*

Name of Features	Mean SHAP Value
page_rank	0.1011105991
google_index	0.08729465626
nb_www	0.07077475209
phish_hints	0.04106446658
nb_hyperlinks	0.0373490615
length_hostname	0.03638746815
nb_hyphens	0.03345303188
longest_words_raw	0.02385161969
nb_slash	0.02108138988
safe_anchor	0.01928737874
nb_dots	0.0176940118
ratio_extHyperlinks	0.01762839899
shortening_service	0.01661600421
domain_age	0.01519877477
ratio_digits_host	0.01506407834
https_token	0.01500930474
domain_in_title	0.01468479694
ip	0.01358343854
domain_in_brand	0.01337623824

Figure 10. Mean SHAP Values for top 20 Features

These values were sorted in descending order to highlight the features with higher influence on predictions.

The comprehensive experimental setup encompassed the construction and training of an ANN model using Keras for URL classification which achieved the accuracy of 95.93% on test dataset. Following the model training, SHAP analysis provided insights into feature importance, aiding in understanding the ANN model's decision-making process for classifying 'phishing' and 'legitimate' URLs.

CONCLUSION

In this study, we explored the efficacy of two distinct machine learning models, Random Forest and Artificial Neural Network (ANN) implemented using Keras, for the crucial task of phishing URL detection. The primary goal was to comprehend the decision-making processes of these models and ascertain the significance of individual features in differentiating between legitimate and malicious URLs. The Random Forest model exhibited commendable performance, achieving an accuracy of approximately 96.59% on a testing set. This model's interpretability was further enhanced through SHAP (SHapley Additive exPlanations) analysis, unveiling feature importance. Analysis revealed that features related to URL characteristics, domain properties, and web traffic statistics held substantial weight in the model's decision-making process. Similarly, the ANN model, with an accuracy of around 95.93%, demonstrated robustness in distinguishing between phishing and legitimate URLs. Employing SHAP analysis shed light on feature relevance, emphasizing the influence of various URL attributes and structural properties in the ANN's predictive outcomes.

REFERENCES

[1] R. Zieni, L. Massari, and M. C. Calzarossa, "Phishing or Not Phishing? A Survey on the Detection of Phishing Websites," in IEEE Access, vol. 11, pp. 18499-18519, 2023, doi: 10.1109/ACCESS.2023.3247135.

[2] P. Deore and N. Mishra, "India Cyber Threat Report 2023," Data Security Council of India, Retrieved January 5, 2024, from https://www.dsci.in.

[3] IBM, "Threat Intelligence," Retrieved January 5, 2024, from https://www.ibm.com/reports/threat-intelligence.

[4] F. Greco, G. Desolda, and A. Esposito, "Explaining Phishing Attacks: An XAI Approach to Enhance User Awareness and Trust," 2023.

[5] M. Vijayalakshmi, S. Mercy Shalinie, M. H. Yang, and R. Meenakshi, "Web phishing detection techniques: a survey on the state-of-the-art, taxonomy, and future directions," IET Networks, vol. 9, no. 5, pp. 235-246, 2020.

[6] S. Sheng et al., "An empirical analysis of phishing blacklists," 2009.

[7] O. K. Sahingoz et al., "Machine learning based phishing detection from URLs," Expert Systems with Applications, vol. 117, pp. 345-357, 2019.

[8] P. Yang, G. Zhao, and P. Zeng, "Phishing website detection based on multidimensional features driven by deep learning," IEEE Access, vol. 7, pp. 15196-15209, 2019.

[9] E. Zhu et al., "DTOF-ANN: An Artificial Neural Network phishing detection model based on Decision Tree and Optimal Features," Applied Soft Computing, vol. 95, 2020.

[10] M. Moghimi and A. Y. Varjani, "New rule-based phishing detection method," Expert Systems with Applications, pp. 231-242, 2016.

[11] C. Y. Wu, C. C. Kuo, and C. S. Yang, "A phishing detection system based on machine learning," in 2019 International Conference on Intelligent Computing and its Emerging Applications (ICEA), pp. 28-32.

[12] R. Alazaidah et al., "Website Phishing Detection Using Machine Learning Techniques," Journal of Statistics Applications & Probability, vol. 13, no. 1, Article 8, 2024.

[13] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017, pp. 4765-4774.

[14] S. Sountharrajan et al., "Wireless Communication in Cyber Security," pg. 160, November 2023.

[15] V. Hassija et al., "Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence," Cognitive Computation, 2023.

[16] W. J. von Eschenbach, "Transparency and the Black Box Problem: Why We Do Not Trust AI," Philosophy & Technology, vol. 34, pp. 1607-1622, 2021.

[17] A. Hannousse and S. Yahiouche, "Web page phishing detection," Mendeley Data, V3, doi: 10.17632/c2gw7fy2j4.3, 2021.

[18] M. HR et al., "Development of anti-phishing browser based on random forest and rule of extraction framework," Cybersecurity, vol. 3, no. 20, 2020.

[19] S. Alnemari and M. Alshammari, "Detecting Phishing Domains Using Machine Learning," Applied Sciences, vol. 13, no. 8, p. 4649, 2023.

[20] J. Patel, "Phishing URL Detection using Artificial Neural Network," IJRESM, vol. 5, no. 4, pp. 47–51, Apr. 2022.

[21] A. A. Alkhouly, A. Mohammed, and H. A. Hefny, "Improving the Performance of Deep Neural Networks Using Two Proposed Activation Functions," IEEE Access, vol. 9, pp. 82249-82271, 2021, doi: 10.1109/ACCESS.2021.3085855.

[22] I. Muraina, "IDEAL DATASET SPLITTING RATIOS IN MACHINE LEARNING ALGORITHMS: GENERAL CONCERNS FOR DATA SCIENTISTS AND DATA ANALYSTS," 2022.

[23] M. Zouina and B. Outtaj, "A novel lightweight URL phishing detection system using SVM and similarity index," Human-centric Computing and Information Sciences, vol. 7, no. 17, 2017.

[24] R. Rodríguez-Pérez and J. Bajorath, "Interpretation of compound activity predictions from complex machine

learning models using local approximations and Shapley values," Journal of Medicinal Chemistry, 2020.

[25] W. E. Marcílio and D. M. Eler, "From explanations to feature selection: assessing SHAP values as feature selection mechanism," in 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Porto de Galinhas, Brazil, pp. 340-347, doi: 10.1109/SIBGRAPI51738.2020.00053.