ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 4, April 2024

PhishGuard: Leveraging Machine Learning For Phishing URL Detection

[1] P. Venkata Jahnavi, [2] P. Hima Sumana, [3] Sk. Charishma Kousar, [4] P. Himaja, [5] K. Jeevan Ratnakar

[1] [2] [3] [4] B.Tech, Students. Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India

[5] Assistant Professor, Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India

Corresponding Author Email: [1] jahnavipunati33@gmail.com, [2] himasumana123@gmail.com, [3] charishmakousar@gmail.com, [4] himajapasupuleti@gmail.com, [5] jeevanratnakarvvit@gmail.com

Abstract*— Phishing is a sort of online fraud in which fraudulent emails and websites deceive victims into disclosing important information. To combat this problem, scientists have developed tactics aimed at creating effective phishing URL detection systems. To this end, our method examines URL attributes such as domain-based and address-based data using feature engineering and ensemble machine learning techniques. This work presents an advanced phishing URL detection system that analyses URL properties using feature engineering and ensemble machine learning approaches. Our approach incorporates the Gradient Boosting Classifier, outperforming other algorithms that mostly use Random Forest, Decision Tree, and Logistic Regression, and achieves an amazing 97.6% accuracy in differentiating between phishing and authentic websites. Phishing is a common online danger that uses shady websites and emails to trick people into disclosing personal information. Our solution guarantees accuracy and efficiency while addressing issues related to computational complexity. We showcase our system's performance using precision, recall, accuracy, and F1 score metrics through a thorough examination of various datasets. Furthermore, our system has an intuitive user interface that combines HTML, CSS, and Flask to improve usability and accessibility. By making a substantial improvement to phishing detection, this study protects consumers from online risks and maintains their privacy.

***Keywords*— Domain Based Feature, Flask, Machine Learning Algorithms, Phishing, URL.**

I. INTRODUCTION

In the rapidly evolving digital world, the need for robust and forward-thinking security measures is paramount, particularly in the face of growing cyber threats such as phishing. Phishing is a form of cyber attack where fraudsters trick individuals into revealing confidential information via deceptive websites, posing a substantial risk to cybersecurity. Conventional methods for identifying phishing URLs often struggle to keep pace with the sophisticated strategies employed by cybercriminals. These attackers craft phishing URLs and use manipulative tactics to fool individuals into clicking the link and inputting personal data.

To safeguard against phishing attacks, users must exercise caution when interacting with links in unexpected communications, meticulously inspect the URL for any anomalies or misspellings, and refrain from providing sensitive information on websites lacking a secure connection (https://). Phishing URLs are typically disseminated through platforms like online banking, social media, or instant messaging apps. Addressing this issue, PhishGuard presents an innovative solution that utilizes sophisticated machine learning algorithms to proactively detect and neutralize phishing threats.

In the face of the ever-changing digital threat landscape, our initiative, PhishGuard, adopts a forward-thinking stance by leveraging machine learning capabilities. By integrating sophisticated algorithms with comprehensive datasets, PhishGuard strives to improve the identification of phishing URLs, thereby providing a solid line of defence against cyber threats. Thanks to the incorporation of a machine learning model, PhishGuard can adapt and evolve in response to emerging phishing strategies.

Among these strategies, gradient boosting has demonstrated its effectiveness in differentiating between legitimate and malicious URLs. Gradient boosting is a type of ensemble learning method that amalgamates the capabilities of multiple weak learners to construct a strong and precise predictive model. The process initiates with the compilation of a diverse array of features extracted from URLs, which include aspects such as URL length, lexical content, domain attributes, and structural patterns.

Fig 1. Components of URL

The phishing detection mechanism utilizes a variety of machine learning models, including a Decision Tree, Support

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 4, April 2024

Vector Machine, XGBoost, Multilayer Perceptron, Gradient Boosting, and Random Forest. The power of gradient boosting lies in its capacity to manage intricate relationships within the data and prevent overfitting, leading to a highly adaptable and accurate phishing URL detection model. The gradient boosting algorithm iteratively constructs a series of decision trees, each focusing on specific facets of the feature space. The model continually refines itself, placing greater emphasis on misclassified instances, thereby enhancing overall accuracy. This approach excels in identifying subtle patterns and anomalies that may evade traditional methods.

Phishing URLs may display complex, non-linear patterns and interactions among features. Gradient boosting algorithms are adept at capturing such non-linear relationships and interactions, making them apt for detecting sophisticated phishing attacks. With its ability for continuous learning and adaptation, gradient boosting serves as a formidable ally in the ongoing fight against cyber threats. Gradient boosting algorithms incorporate mechanisms, such as regularization techniques, that aid in preventing overfitting. This is vital for phishing URL detection, as overfitting can result in poor generalization on unseen data. Gradient boosting models offer insights into feature importance, assisting analysts and cybersecurity experts in understanding which features contribute most to the classification decision. This interpretability is invaluable for pinpointing the characteristics of phishing URLs.

Fig 2. Difference between Legitimate and Phishing URL

Our initiative prioritizes not only accuracy in identifying malicious URLs but also emphasizes efficiency to minimize false positives and ensure a smooth user experience. Machine learning algorithms can be employed to examine a URL's path, domain, and parameters, among other features, to identify phishing URLs. The current phishing detection system employs supervised machine learning techniques, one of which is text categorization. Text preprocessing, text representation for feature extraction, and classification constitute the three primary phases of this system. The existing system leverages handcrafted features for detection, which are manually designed and engineered attributes derived from analyzing specific characteristics of URLs, such as symbol frequency, domain length, special symbols, and the presence of common top-level domains. A machine learning-based model has been developed for the proposed system to recognize phishing and authentic URLs. It follows two main steps: first, searching the URL, and second, extracting domain-based features for analysis. The phishing dataset includes various features such as "Age of the Domain and Subdomain", "Prefix and Suffix", "Google Index", "HTTPS", "AnchorURL", and "WebsiteTraffic" and uses F1 Score and Recall as performance metrics.

II. LITERATURE REVIEW

Phishing is a significant cyber threat that involves tricking users into providing credentials through counterfeit login forms. This method proposes to detect phishing websites by analyzing URLs and comparing deep learning and machine learning methods. Unlike current methods, which often overlook login sites, we include them in both phishing and authentic classes to provide a more accurate assessment. It shows that current methods have substantial false-positive rates when tested with real login URLs. Additionally, it examines the temporal component of model correctness and conducts a frequency study of phishing domains. To support our findings, we present the Phishing Index Login URL (PILU-90K) dataset, which consists of 30K phishing URLs and 60K legitimate URLs. Finally, a Logistic Regression model with TF-IDF feature extraction achieves 96.50% accuracy on the login URL dataset [1].

Phishing is a prevalent form of cybercrime where cybercriminals use deceptive methods to trick users into revealing sensitive information by creating emails or websites that appear genuine. In response to this problem, researchers and practitioners have developed various techniques to recognize phishing URLs. The aim is to assist in this endeavour by developing an advanced system for identifying phishing websites. The system will examine several URL attributes, including the domain name, path, length, and presence of dubious keywords, using feature engineering and machine learning methods. Metrics including precision, recall, accuracy, and F1 score will be used to evaluate performance, and a large dataset will be used for the evaluation. To assess the system's effectiveness and efficiency, it will also be compared to existing technologies in the market [2].

Phishing is a persistent social engineering crime that still poses a serious risk, especially to the financial industry. Despite much research, robust and long-lasting methods to counteract phishing attempts are still unclear. This is a novel approach to phishing detection that combines the concepts of Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) architectures, utilizing both URLs and HTML pages. The LSTM network is used in conjunction with 1D convolutional layers to learn URL features, while an alternative 1D convolutional network is used to extract

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 4, April 2024

features from HTML text. The suggested model is the result of these two networks being independently trained and then merged together using a sigmoid layer. With an accuracy of 98.34%, the suggested model outperforms the highest accuracy of 97.3%. Additionally, the model's feature extraction does not rely on external services, which makes it easier to create effective real-time phishing detection tools that protect Internet users [3].

This Systematic Literature Review (SLR) thoroughly examines and compares several phishing detection strategies, including Lists Based, Visual Similarity, Heuristic, Machine Learning, and Deep Learning. The SLR examines the algorithms and methods used in the detection of phishing websites and offers research questions that could be pursued further. This research constitutes a revision of earlier systematic literature reviews, emphasizing the most recent advancements in phishing detection methods. Through a thorough examination of various approaches, datasets, and performance measures, this article improves readers' understanding of techniques for detecting phishing websites. Notably, the majority of studies (57) use machine learning approaches. Additionally, the poll identifies the websites of PhishTank and Alexa as the main resources for phishing and authentic datasets, respectively. One of the most popular machine learning techniques, the Random Forest Classifier, is used in 31 publications. Convolutional Neural Network (CNN) is the most effective approach, according to several studies, detecting phishing websites with an amazing 99.98% accuracy [4].

Phishing has always posed a serious problem. However new developments in phishing detection, especially those based on machine learning, have shown promise in reducing these attempts. To detect phishing domains, this research develops and compares four machine learning algorithms. Moreover, the most accurate model among the four is compared with an existing solution found in the literature. Artificial neural networks (ANNs), decision trees (DTs), support vector machines (SVMs), and random forests (RF) are all used in the models. The UCI phishing domains dataset serves as an assessment tool, providing a benchmark for the model's performance. Our results show that the random forest-based model outperforms the other three methods and outperforms previously published solutions in the literature in terms of accuracy [5].

III. METHODOLOGY

a. Data Collection

A list of the URLs to over 11,000 websites. Each example consists of thirty parameters for the website along with a class label indicating if the website is a phishing site (1 or -1). The dataset in a ".txt" file is made up entirely of the column values; headers are not present. The column names were appended after the ".csv" file was created.

-1: This value could represent a negative or harmful characteristic. For example, in the context of phishing, it might indicate a suspicious or malicious attribute of a website or email.

0: This value might represent a neutral or inconclusive characteristic. It could signify that there is no clear indication of either phishing or legitimate behavior.

1: Conversely, this value could represent a positive or legitimate characteristic. In the context of phishing, it might indicate traits commonly found in legitimate websites or communications.

- UsingIP (categorical - signed numeric) : { -1,1 }
- LongURL (categorical - signed numeric) : { 1,0,-1 }
- ShortURL (categorical - signed numeric) : { 1,-1 }
- Symbol@ (categorical - signed numeric) : { 1,-1 }
- Redirecting// (categorical - signed numeric) : { -1,1 }
- PrefixSuffix- (categorical - signed numeric) : { -1,1 }
- SubDomains (categorical - signed numeric) : { -1,0,1 }
- HTTPS (categorical - signed numeric) : { -1,1,0 }
- DomainRegLen (categorical - signed numeric) :
- { -1,1 }
- Favicon (categorical - signed numeric) : { 1,-1 }
- NonStdPort (categorical - signed numeric) : { 1,-1 }
- HTTPSDomainURL (categorical - signed numeric) : { -1,1 }
- RequestURL (categorical - signed numeric) : { 1,-1 }
- AnchorURL (categorical - signed numeric) : { -1,0,1 }

b. Data Visualization

Finding patterns, trends, and outliers in data can be made easier with the help of data visualization. The data was previewed using the Python charting libraries Matplotlib and Seaborn for data visualization. We have used a variety of data visualization techniques in our study article to improve our comprehension of phishing and authentic websites. First, we've used a correlation heatmap to show the connections between various website elements. Stronger correlations are indicated by darker shades on the heatmap, which shows correlation coefficients between pairs of variables. With their labels on both axes, features like "URL Length," "having SubDomain," and "SSL final State" reveal information about how they are related to one another. In order to help with pattern and outlier identification, we have also used pair plots to display pairwise correlations and distributions of particular traits. In addition, a pie chart that shows the percentage of each category statistically contrasts the frequency of phishing with occurrences that are valid in our dataset.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 4, April 2024

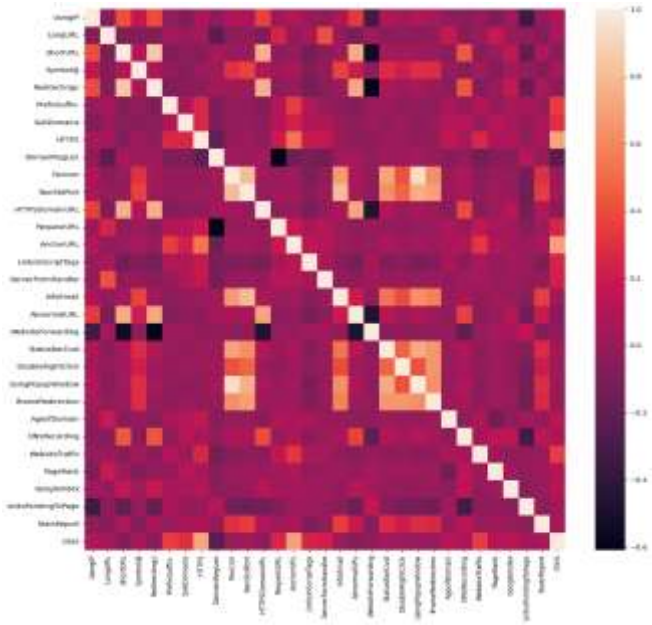

Fig 3. Correlation Heatmap for the dataset Visualization

c. Feature Extraction

In our phishing URL detection project, we've implemented a comprehensive feature extraction mechanism to analyze various attributes of URLs and distinguish between legitimate and phishing websites. This feature extraction process involves examining elements such as short URLs or long, the presence of special symbols, Anchor URLs, redirection patterns, and characteristics of the domain name. Additionally, we delve into the structure of the URL, assessing the presence of subdomains and whether the URL employs standard HTTP or the more secure HTTPS protocol. These features provide valuable insights into the potential malicious intent behind a given URL.

Furthermore, our feature extraction mechanism extends to examining the behaviour of the website itself, including the presence of embedded links, scripts, and form actions. Examining these elements closely allows us to spot oddities like strange redirects, dubious server answers, and possible phishing strategies like email harvesting via embedded scripts. We also assess how well the website uses status bar customisation, a common ruse employed by phishing websites to deceive users. By using this comprehensive feature extraction technique, we hope to develop a robust detection system that can accurately identify phishing URLs and alert users to potential risks before they become the targets of fraudulent.

All things considered, our feature extraction methodology captures a broad spectrum of traits and actions linked to URLs and website architectures, allowing us to create an advanced phishing detection system. By leveraging machine learning algorithms and statistical analysis on these extracted features, we can effectively classify URLs as either legitimate or phishing with high accuracy. By taking a proactive stance in spotting such dangers, users are better equipped to surf the internet safely, reduce their vulnerability to phishing scams, and protect their personal data from unscrupulous parties.

d. Classification Models

Supervised learning, a prevalent and successful machine learning type, allows us to predict specific outcomes or labels based on a set of features. We use examples of feature-label pairs to create our training set and build a machine-learning model. The aim is to make accurate predictions for new, unseen data.

Supervised machine learning problems are primarily divided into two categories: classification and regression. In our case, we're dealing with a regression problem as the predicted suicide rate is a continuous number, also known as a floating-point number in programming terms. We trained our dataset using various regression models for supervised machine learning, including k-nearest Neighbors, Support Vector Classifier, Decision Tree, Random Forest, Gradient Boosting, and Xgboost. The model's performance was evaluated using metrics such as accuracy and F1 score.

In our evaluation of various machine learning models for phishing URL detection, the Gradient Boosting Classifier outperformed its competitors across key metrics. We conducted an extensive analysis to assess the effectiveness of each model, considering accuracy, f1_score, recall, and precision. While models like Decision Tree, K-Nearest Neighbors, and Logistic Regression performed well, the Gradient Boosting Classifier consistently delivered superior results. With an accuracy of 97.4%, f1_score of 0.977, recall of 0.994, and precision of 0.986, the Gradient Boosting Classifier emerged as the top performer. Its ability to accurately distinguish between phishing and legitimate URLs makes it the optimal choice for our research project, ensuring enhanced accuracy and effectiveness in detecting malicious online activities.

	ML Model	Accuracy	f1_score	Recall	Precision
0	Logistic Regression	0.934	0.941	0.943	0.927
1	K-Nearest Neighbors	0.956	0.961	0.991	0.989
2	Support Vector Machine	0.964	0.968	0.980	0.965
3	Decision Tree	0.957	0.961	0.991	0.993
4	Random Forest	0.967	0.971	0.993	0.989
5	Gradient Boosting Classifier	0.974	0.977	0.994	0.986
6	XGBoost Classifier	0.956	0.965	1.000	1.000

Fig 4. Comparison of Different Models Accuracy and Performance Metrics

IV. SYSTEM ARCHITECTURE

An efficient method for detecting phishing attempts and guaranteeing online security is orchestrated by the

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 4, April 2024

architecture shown in the picture. The process begins when a user enters a URL to begin feature extraction. In this stage, relevant URL elements are carefully collected to capture important details, like length, subdomain presence, and SSL status. The extracted features are then analysed using an extensive training dataset that includes labelled instances of URLs that have been classified as either legitimate or phishing.

The classifier model, the central component of the design, finds patterns and establishes a decision limit by applying a range of machine learning techniques, such as Gradient Boosting, Random Forests, Decision Trees, and Logistic Regression. By utilizing the knowledge acquired from the training dataset, the model becomes proficient in distinguishing between phishing and authentic URLs. Ultimately, the architecture's output provides a definitive judgment, classifying the input URL as "Legitimate" or "Phishing" by the attributes and patterns that were learned.

Essentially, this architecture combines machine learning methods, labelled data, and feature extraction to provide a strong mechanism for URL classification. It is essential for protecting consumers from phishing scams and maintaining online security standards since it analyzes URLs methodically and makes use of machine learning.

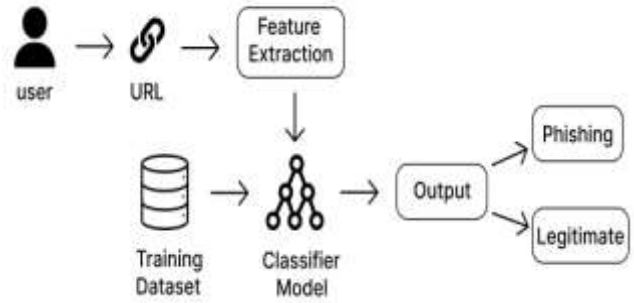

Fig 5. System Architecture of Phishing URL Detection

V. RESULTS

This is a bar plot displaying the results of our thorough comparison of various classification methods, which comprised XGBoost, K-Nearest Neighbors, Random Forest, Decision Tree, Logistic Regression, and Support Vector Machine (SVM). Notably, out of all the options examined, Gradient Boosting turned out to be the most accurate algorithm.

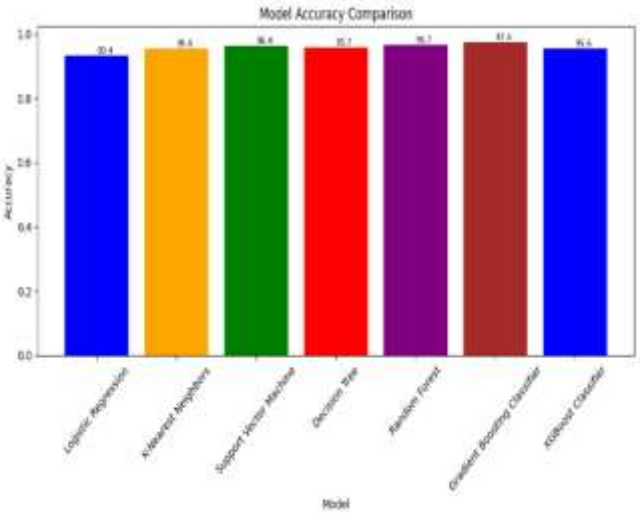

Fig 6. Comparison of Algorithms

In our research, the Gradient Boosting Classifier emerged as a standout performer in distinguishing between phishing and legitimate URLs, showcasing exceptional accuracy and robustness. Throughout the training process, the model demonstrated perfect alignment with ground truth labels, achieving an F1 score of 1.0 and perfect accuracy. Even during testing, the classifier maintained high accuracy and F1 score, indicating effective generalization to unseen data. Furthermore, with recall and precision scores close to 1.0 and approximately 0.972 respectively, the model exhibited proficiency in detecting phishing threats while minimizing false positives. These impressive metrics have significant real-world implications, offering enhanced threat detection capabilities for cybersecurity practitioners and organizations. By deploying the Gradient Boosting Classifier in various security applications, we can protect users from phishing scams, data breaches, and financial losses, thus contributing to a safer digital environment. As we continue our research, further optimization of the model's hyperparameters will continue to strengthen our defenses against evolving cyber threats.

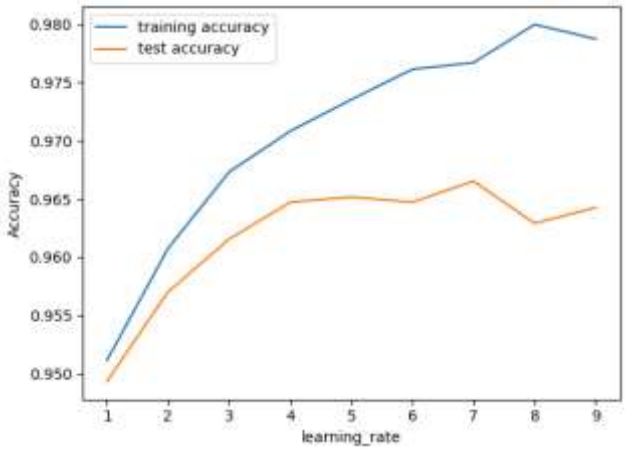

Fig 7. Plotting of training and testing accuracy for n_estimators

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 4, April 2024

In addition to evaluating the Gradient Boosting Classifier's performance through standard metrics, our research also delved into the insights provided by the confusion matrix. This matrix decomposes the model's predictions into true positives, true negatives, false positives, and false negatives. True positives signify instances where the model correctly identified phishing URLs, contributing to effective threat detection. True negatives represent correct identifications of legitimate URLs, enhancing user confidence in the classifier's ability to discern benign content. False positives and false negatives, on the other hand, highlight areas of concern, indicating instances where the model misclassified URLs. We can learn a lot about the model's advantages and disadvantages by examining the confusion matrix. This helps us make improvements to the model that will increase its precision and dependability in spotting phishing threats.

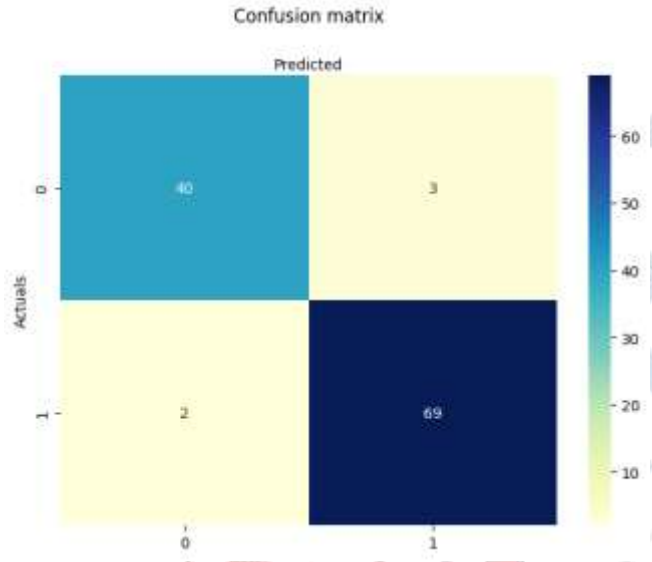

Fig 8. Confusion Matrix For a Binary Classifier

In addition to our model development and evaluation, we have successfully integrated a user interface using Flask to facilitate URL phishing detection. This user interface serves as a practical tool for users to input URLs and receive real-time outputs regarding their phishing status. Leveraging Flask's capabilities, we have created a seamless and intuitive interface that enhances accessibility and usability for end-users. With this integration, users can conveniently assess the security status of URLs, empowering them to make informed decisions while navigating the digital landscape. This practical implementation underscores the real-world applicability and effectiveness of our research findings in combating phishing threats.

Fig 9. Output of Legitimate URL

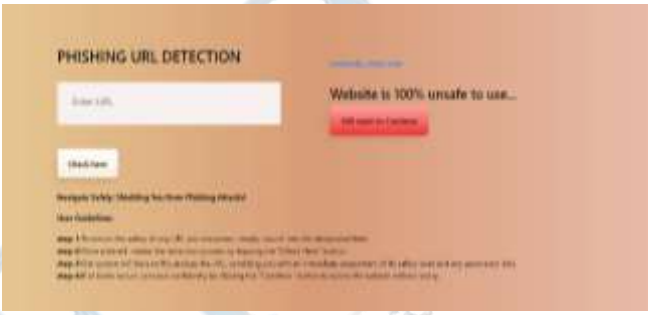

Fig 10. Output Of Phishing URL

VI. CONCLUSION

Finally, our study has thoroughly examined domain-based attributes and used them with an advanced classifier model to effectively counteract phishing attacks. Using a thorough analysis and feature extraction process, we were able to determine important features from URLs, including but not limited to domain length, the existence of subdomains, SSL status, and more, to identify patterns suggestive of phishing efforts. The Gradient Boosting Classifier outperformed the other machine-learning models in our study, showing exceptional accuracy, recall, and precision in differentiating between phishing and authentic URLs. In addition, the incorporation of domain-based characteristics into the classifier model enabled strong threat identification, as demonstrated by the understanding obtained from the confusion matrix. Furthermore, the creation of an intuitive user interface with Flask improves accessibility and gives users the ability to decide in real-time if a URL is secure or not. All things considered, our findings highlight the effectiveness of combining domain-specific features with cutting-edge machine learning methods, providing viable ways to counteract phishing schemes and strengthen cybersecurity protocols in the digital sphere. We want to significantly improve our capacity to prevent emerging cyber threats and guarantee a safer online experience for every user as we continue to hone and optimize our strategy.

VII. FUTURE SCOPE

Our anti-phishing solution will require additional work, in addition to our existing research endeavours, to be optimized for deployment on the Internet of Things (IoT) and mobile devices. To effectively prevent phishing attacks, many

systems have special characteristics and limitations that call for customized solutions. We can expand our system's functionality and offer complete security in a variety of digital scenarios by adding support for mobile and Internet of Things devices. In addition, as part of our continuous development, we want to improve the system's ability to identify links that can only be clicked once. These linkages present a major challenge to existing models since they are frequently ephemeral and intended to avoid typical detection techniques.

REFERENCES

[1] Sánchez-Paniagua, M., Fernández, E. F., Alegre, E., Al-Nabki, W., & Gonzalez-Castro, V. (2022). Phishing URL detection: A real-case scenario through login URLs. IEEE Access, 10, 42949-42960.

[2] James, J., Sandhya, L., & Thomas, C. (2013, December). Detection of phishing URLs using machine learning techniques. In 2013 international conference on control communication and computing (ICCC) (pp. 304-309). IEEE.

[3] Ariyadasa, S., Fernando, S., & Fernando, S. (2020). Detecting phishing attacks using a combined model of LSTM and CNN. Int. J. Adv. Appl. Sci, 7(7), 56-67.

[4] Safi, A., & Singh, S. (2023). A systematic literature review on phishing website detection techniques. Journal of King Saud University-Computer and Information Sciences.

[5] Alnemari, S., & Alshammari, M. (2023). Detecting phishing domains using machine learning. Applied Sciences, 13(8), 4649.

[6] Joshi, Y., Saklikar, S., Das, D., & Saha, S. (2008, December). PhishGuard: a browser plug-in for protection from phishing. In 2008 2nd International Conference on Internet Multimedia Services Architecture and Applications (pp. 1-6). IEEE.

[7] Chen, Juan, and Chuanxiong Guo. "Online detection and prevention of phishing attacks." In 2006 First International Conference on Communications and Networking in China, pp. 1-7. IEEE, 2006.

[8] Pandey, M. K., Singh, M. K., Pal, S., & Tiwari, B. B. (2023). Prediction of phishing websites using machine learning. Spatial Information Research, 31(2), 157-166.

[9] Lakshmi, V. S., & Vijaya, M. S. (2012). Efficient prediction of phishing websites using supervised learning algorithms. Procedia Engineering, 30, 798-805.

[10] Aburrous, M., Hossain, M. A., Dahal, K., & Thabtah, F. (2010, April). Predicting phishing websites using classification mining techniques with experimental case studies. In 2010 seventh international conference on information technology: New generations (pp. 176-181). IEEE.

[11] Preeti, Nandal, R., & Joshi, K. (2021). Phishing URL Detection Using Machine Learning. In Advances in Communication and Computational Technology: Select Proceedings of ICACCT 2019 (pp. 547-560). Springer Singapore.

[12] Ahammad, S. H., Kale, S. D., Upadhye, G. D., Pande, S. D., Babu, E. V., Dhumane, A. V., & Bahadur, M. D. K. J. (2022). Phishing URL detection using machine learning methods. Advances in Engineering Software, 173, 103288.

[13] Oram, E., Dash, P. B., Naik, B., Nayak, J., Vimal, S., & Nataraj, S. K. (2021). Light gradient boosting machine-based phishing webpage detection model using phisher website features of mimic URLs. Pattern Recognition Letters, 152, 100-106.

[14] Abdul Samad, S. R., Balasubaramanian, S., Al-Kaabi, A. S., Sharma, B., Chowdhury, S., Mehbodniya, A., ... & Bostani, A. (2023). Analysis of the performance impact of fine-tuned machine learning model for phishing URL detection. Electronics, 12(7), 1642.

[15] Madhu Chandra, S., & Chandrashekar, K. T. (2020). Malicious url detection using extreme gradient boosting technique. International Research Journal of Modernization in Engineering Technology and Science, 2(10), 675-682.

[16] Karim, A., Shahroz, M., Mustofa, K., Belhaouari, S. B., & Joga, S. R. K. (2023). Phishing detection system through hybrid machine learning based on URL. IEEE Access, 11, 36805-36822.