INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 609-611

IDENTIFYING PHISHING WEBSITES THROUGH MACHINE LEARNING

Dr. Balakrishnan K[1], Srujan M S Aathreya[2], Sourav Sajesh[3], Vishnu T A[4], Chandan Kumar K R[5]

[1]Associate Professor, Dept. Of CSE, Sambhram Institute of Technology, Bengaluru Urban, Karnataka, India.

[2,3,4,5]Final Year Students, Dept. Of CSE, Sambhram Institute of Technology, Bengaluru Urban, Karnataka, India.

ABSTRACT

In recent times, as mobile device usage has surged, there's been a notable shift towards conducting a wide array of real-world activities in the digital realm. While this transition has undeniably added convenience to our daily routines, it has also opened up avenues for security breaches owing to the inherently anonymous nature of the internet. Although antivirus programs and firewall systems offer a degree of protection against such threats, adept attackers often exploit vulnerabilities in users' behavior, employing tactics like phishing through counterfeit websites. These deceptive pages mimic popular platforms such as banking, social media, and e-commerce sites, aiming to pilfer sensitive information like login credentials, financial details, and credit card numbers. Detecting phishing attempts poses a significant challenge, and various approaches have been proposed, including blacklisting, rule-based detection, and anomaly-based detection. Notably, there's a growing focus in the research community on leveraging machine learning-based anomaly detection for its adaptability, particularly in identifying "zero-day" attacks.

Keywords: Mobile device usage, Digital Realm, Security Breaches, Antivirus programs, Phishing.

1. INTRODUCTION

In our daily lives, much of our activities now occur in the digital realm. The use of computers and the internet has become integral to various aspects of both our personal and professional spheres, enabling swift completion of tasks across diverse sectors such as commerce, healthcare, education, communication, finance, aviation, research, engineering, entertainment, and public services. Thanks to advancements in mobile and wireless technologies, accessing the internet from anywhere at any time has become effortless for users requiring connection to local networks. However, this widespread digital integration has brought to light significant vulnerabilities concerning information security. Consequently, there's a pressing need for individuals navigating cyberspace to adopt precautionary measures against potential cyber threats. These threats can originate from various sources, including cybercriminals, hackers, non-malicious actors (often termed "white-hat" attackers), and hacktivists. Their objectives may range from gaining unauthorized access to computers or sensitive information to perpetrating fraud, forgery, extortion, hacking, denial of service attacks, dissemination of malware, distribution of illegal digital content, or employing social engineering tactics. These attacks, spanning from historical instances like the 1988 Morris Worm to contemporary incidents, are typically aimed at defrauding, disrupting, or manipulating target users to obtain valuable information or financial gain.

2. METHODOLOGY

A Phishing Website Dataset has been acquired and processed using various machine learning techniques. The data undergoes preprocessing and is then split into training and testing sets. A prediction model is constructed employing machine learning algorithms such as Logistic Regression, KNN, SVC, Random Forest, Decision Tree, Naïve Bayes, and XGB Classifier. The model is trained with the training dataset and subsequently tested with the testing dataset to evaluate accuracy. The algorithm yielding the highest accuracy is selected as the final prediction model. This finalized model is serialized into a pickle file (binary format) for storage. A user-friendly Front End is developed utilizing Flask and HTML. Users input website links into the Front End, whose parameters are then fed into our finalized algorithm to determine whether the entered website is a phishing site. The predicted outcome is then displayed on the Front End interface.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 609-611

3. MODELING AND ANALYSIS

3.1 System Design:

3.2 Data Flow Diagram

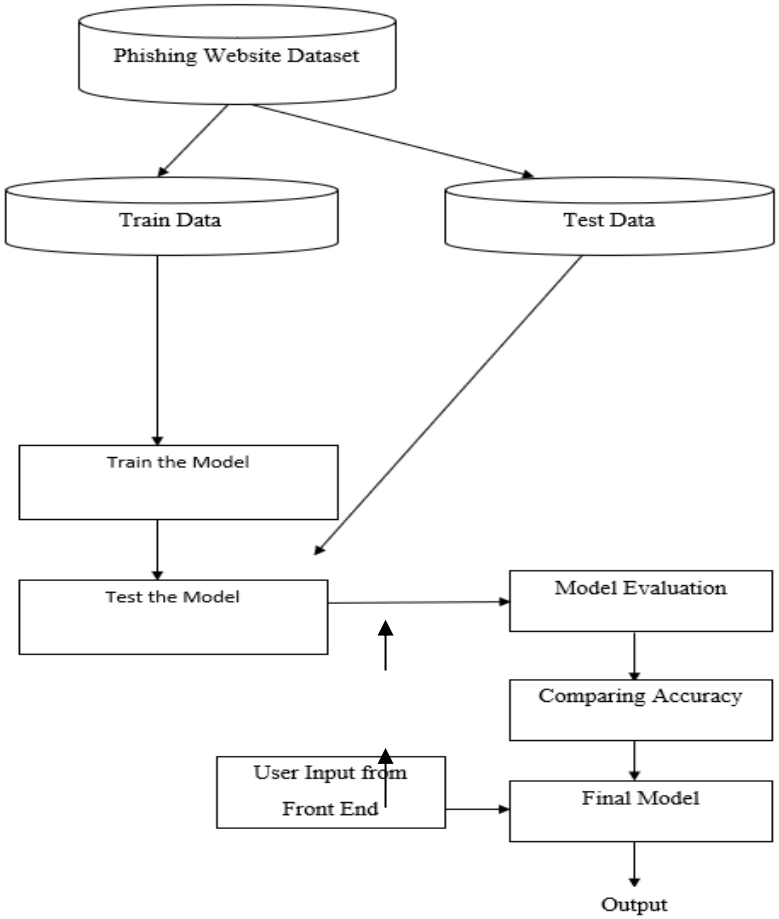

Figure 3.1: System Architecture

4. RESULTS AND DISCUSSION

Phishing is a common attack used to obtain sensitive information using visually similar websites to that of legitimate websites. With the growing technology, phishing attacks are on the rise. Machine Learning is a very popular approach to detect phishing websites.

So, we proposed a system with the help of machine learning techniques and algorithms like Logistic Regression, KNN, SVC, Random Forest, Decision Tree, XGB Classifier and Naïve Bayes to predict Phishing Website based on different parameters like extracted by the website link entered by the user in the front end.

We implemented our project and the system predicts Phishing Website with good accuracy of 96% given by Random Forest Classifier.

Sl.No	Models	Accuracy
1	Naïve Bayes	65.56%
2	Logistic Regression	93.68%
3	Random Forest	96.00%
4	KNN	92.99%
5	Decision Tree	93.68%
6	SVC	94.79%
7	XGB Classifier	95.30%

5. CONCLUSION

In conclusion, machine learning proves effective in detecting phishing websites. Various algorithms are evaluated, with the most accurate selected for deployment. The model is integrated into a user-friendly Front End, empowering users to identify potential threats and bolster their online security.

6. REFERENCES

[1] J. Rashid, T. Mahmood, M. W. Nisar and T. Nazir, "Phishing Detection Using Machine Learning Technique," 2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH), 2020, pp. 43-46, doi: 10.1109/SMART-TECH49988.2020.00026.

[2] M. H. Alkawaz, S. J. Steven and A. I. Hajamydeen, "Detecting Phishing Website Using Machine Learning," 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA), 2020, pp. 111-114, doi: 10.1109/CSPA48992.2020.9068728.

[3] V. Patil, P. Thakkar, C. Shah, T. Bhat and S. P. Godse, "Detection and Prevention of Phishing Websites Using Machine Learning Approach," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1-5, doi: 10.1109/ICCUBEA.2018.8697412.

[4] W. Bai, "Phishing Website Detection Based on Machine Learning Algorithm," 2020 International Conference on Computing and Data Science (CDS), 2020, pp. 293-298, doi: 10.1109/CDS49703.2020.00064.