e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 7, July 2024

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.521

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707131 |

Phising Websites Features Classification based on Extreme Machine Learning

Prof. Barnali Chakraborty, Fanish Kumar Choudhary, Dr. Pavan. G P

Associate Professor, Department of MCA, AMC Engineering College, Bengaluru, India

4th Semester MCA, Department of MCA, AMC Engineering College, Bengaluru, India

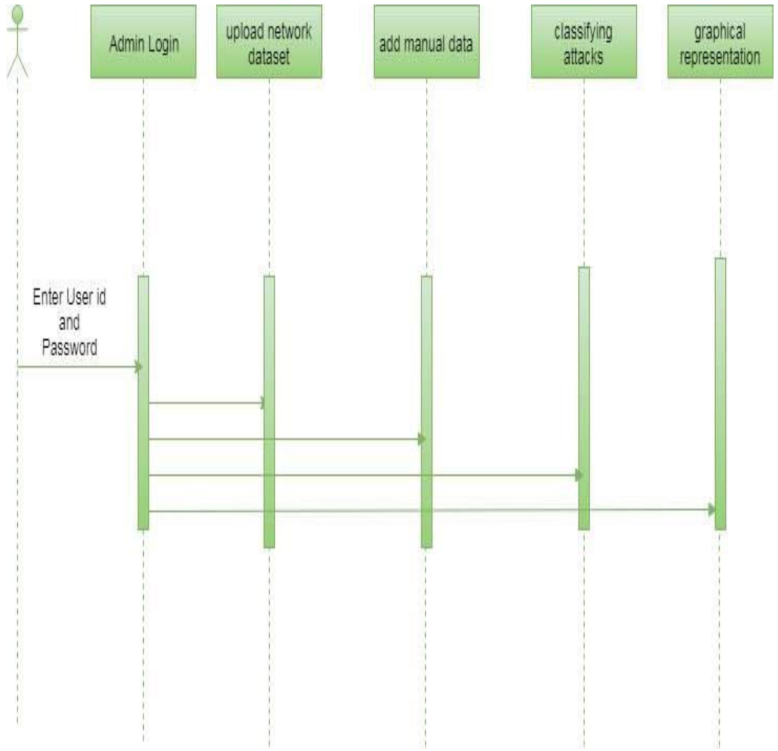
Department of ISE, AMC Engineering College, Bengaluru, India

ABSTRACT: A phishing website is a fraudulent website designed to trick users into believing it is a legitimate website. They often copy the look and feel of trusted websites such as banks, social media platforms or online stores. The goal is to trick visitors into entering sensitive information that an attacker could use for malicious purposes, such as access to credentials or financial information.

KEYWORDS: Fraud, deception, imitation, sensitive information, evidence, money, bad faith.

I. INTRODUCTION

Phishing websites are fake online platforms designed to look like legitimate websites. They usually follow the opinions of reliable sources such as banks, social media platforms or online stores. The main purpose of these scam websites is to trick users into revealing sensitive information such as passwords, credit card numbers or personal information. This information is used by cybercriminals to commit various crimes and poses a serious threat to network security.



II. PHIAHING WEBSITE FEATURES TECHNOLOGIES AND STANDARDS

The following section provides a brief overview of the norms used for Phishing website features.
Phishing websites pose a significant threat to cybersecurity, and using machine learning to detect these threats is an effective strategy. Here's an overview of how machine learning technologies and human-AI language standards can be leveraged to identify phishing websites:

1. Features of Phishing Websites

- **URL Characteristics**:
  - **Length**: Phishing URLs tend to be longer and more complex.
  - **Domain Age**: Many phishing websites are newly created.
  - **HTTPS**: Absence of HTTPS can be a red flag, though not always.
  - **Special Characters**: Presence of unusual characters or excessive subdomains.
- **Content-Based Features**:
  - **Keywords**: Use of urgent or alarming language (e.g., "urgent," "immediate action required").
  - **Visual Similarity**: Mimicking the appearance of legitimate websites.
  - **Login Forms**: Presence of login forms on non-login pages.
- **Technical Indicators**:
  - **IP Address**: Direct IP addresses rather than domain names.
  - **Server Location**: Inconsistencies in server locations.
- **Behavioral Features**:
  - **Traffic Patterns**: Unusual traffic spikes.
  - **User Behavior**: Analysis of user interaction patterns.

2. Machine Learning Techniques

- **Supervised Learning**:
  - **Classification Algorithms**: Decision Trees, Random Forest, Support Vector Machines (SVM), Neural Networks.
  - **Training Data**: Labeled datasets with features of phishing and legitimate websites.
  - **Feature Selection**: Identifying and selecting the most relevant features.
- **Unsupervised Learning**:
  - **Clustering**: Identifying patterns and grouping similar phishing websites.
  - **Anomaly Detection**: Detecting outliers that deviate from normal behavior.
- **Natural Language Processing (NLP)**:
  - **Text Analysis**: Analyzing the language used in website content and emails.
  - **Sentiment Analysis**: Understanding the tone and urgency of the content.
- **Ensemble Methods**:
  - Combining multiple machine learning models to improve accuracy.

III. METHODOLOGY

Phishing websites pose significant threats by masquerading as legitimate sites to deceive users into providing sensitive information. Machine learning (ML) methodologies have emerged as effective tools for detecting such phishing attempts by analyzing various features of URLs and web pages. Below is a detailed overview of the features and methodologies used in phishing website detection through machine learning.

The detection of phishing websites typically involves extracting a variety of features from both phishing and legitimate URLs. Common categories of features include:

1. Address Bar Features

Length of URL: Phishing URLs are often longer than legitimate ones.
Presence of HTTPS: Legitimate sites usually have HTTPS, while many phishing sites do not.
Use of IP Address: Phishing sites may use an IP address instead of a domain name.

2. **Domain Features**
   Domain Age: New domains are often used for phishing.
   Domain Registration Details: Information such as WHOIS data can indicate legitimacy.

3. **HTML and JavaScript Features**
   Presence of Forms: Phishing sites often contain forms to capture user data.
   JavaScript Usage: Excessive or suspicious JavaScript can indicate phishing.

4. **Content-Based Features**
   Text and Image Analysis: Analyzing the content of the webpage for similarities to known legitimate sites.

5. **Behavioral Features**
   User Interaction Patterns: Monitoring how users interact with the site can reveal phishing attempts.
   Machine Learning Methodologies

Various machine-learning techniques are employed to classify URLs as phishing or legitimate. These include:

Supervised Learning

Classification Algorithms: Models such as Decision Trees, Random Forests, Support Vector Machines (SVM), and Neural Networks are trained on labeled datasets to classify URLs.

Unsupervised Learning

Anomaly Detection: This method identifies patterns and anomalies in unlabeled data, which may indicate phishing.

Deep Learning

Convolutional Neural Networks (CNNs): Used for analyzing images and complex patterns in web pages.
Recurrent Neural Networks (RNNs): Effective in processing sequences, such as user interaction logs.

Ensemble Learning

Combining Models: Techniques that combine multiple models to improve detection accuracy and robustness.

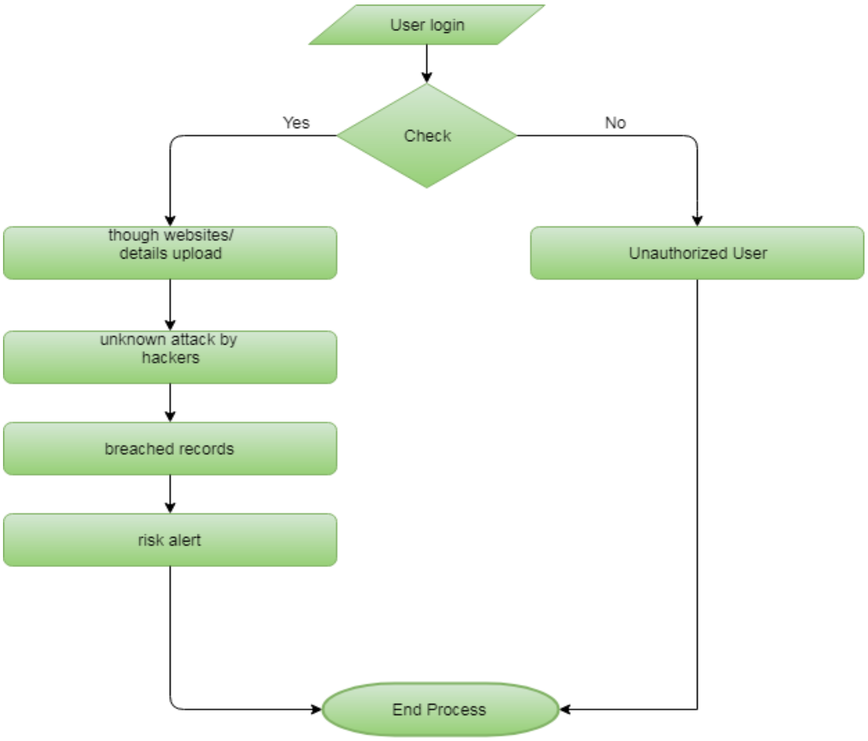


Figure 2. Data Flow Diagram

IV. EXPERIMENTAL RESULTS

1. LANDING PAGE

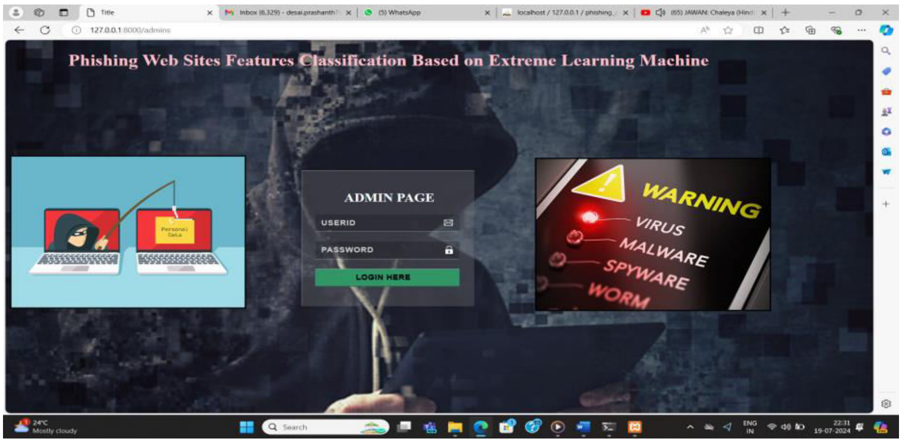

Phishing Web Sites Features Classification Based on Extreme Learning Machine
ADMIN PAGE
USERID
PASSWORD
LOGIN HERE
WARNING
VIRUS
MALWARE
SPYWARE
WORM

Figure 3. LANDING PAGE

2. ADMIN PAGE

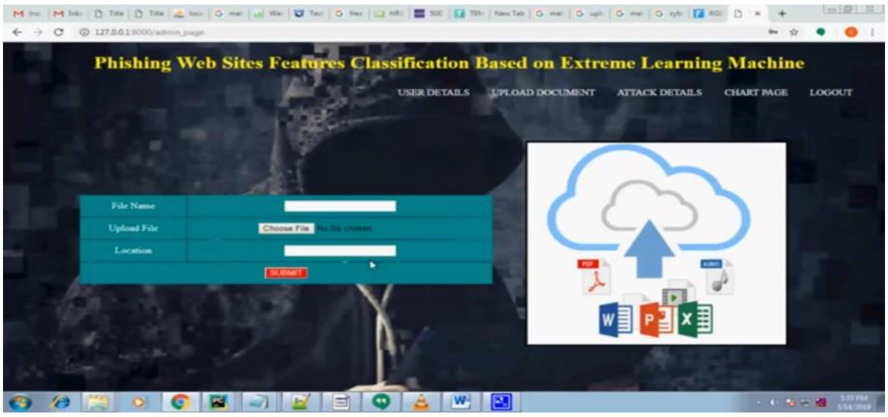


Figure 4. ADMIN PAGE

3. ATTACK INFORMATION

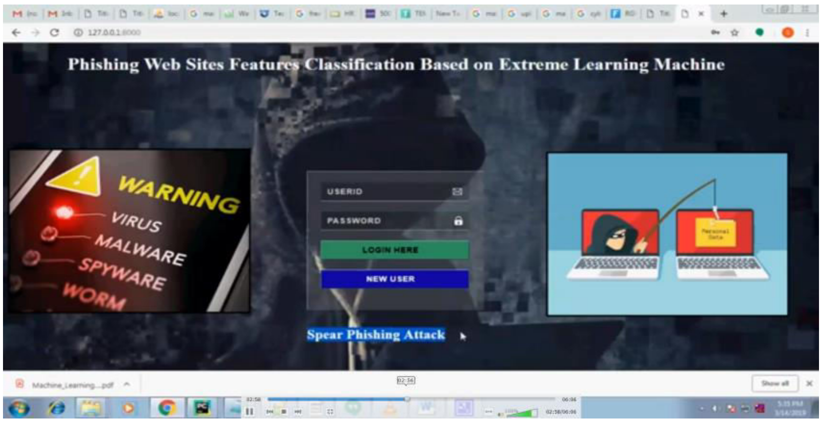


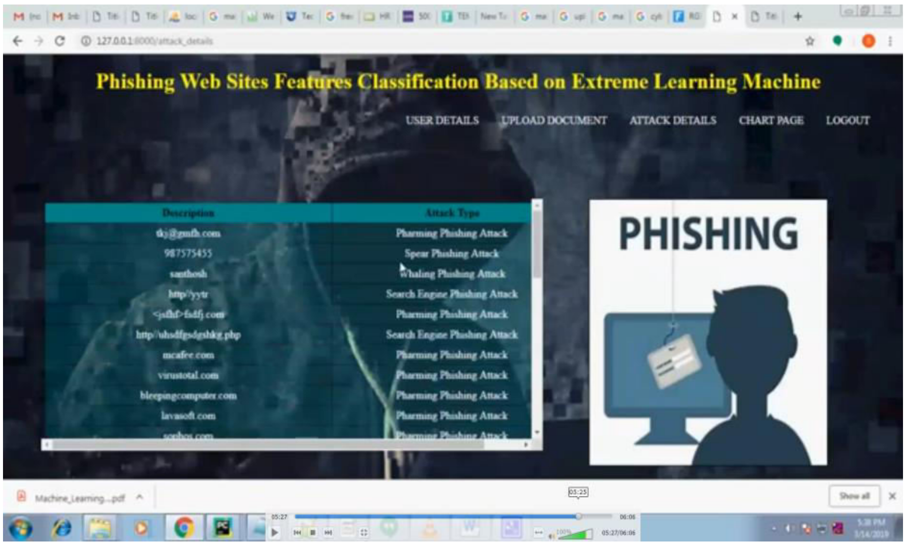
Figure 5. ATTACK INFORMATION

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707131 |

4. DETAILS OF ATTACK


127.0.0.1:8000/attack_details
Phishing Web Sites Features Classification Based on Extreme Learning Machine
USER DETAILS
UPLOAD DOCUMENT
ATTACK DETAILS
CHART PAGE
LOGOUT
PHISHING
Machine_Learning....pdf
Show all

Description	Attack Type
tkj@gmfh.com	Pharming Phishing Attack
987575455	Spear Phishing Attack
santhosh	Whaling Phishing Attack
http//yytr	Search Engine Phishing Attack
<jsfhf>fsdfj.com	Pharming Phishing Attack
http//uhsdfgsdgshkg.php	Search Engine Phishing Attack
mcafee.com	Pharming Phishing Attack
virustotal.com	Pharming Phishing Attack
bleepingcomputer.com	Pharming Phishing Attack
lavasoft.com	Pharming Phishing Attack
sophos.com	Pharming Phishing Attack

Figure 6. DETAILS OF ATTACK

V. CONCLUSION

A categorization model for phishing assaults was developed in this work, and the aspects of the attacks were outlined in this study, as well. Features extracted from websites and a categorization section are part of this method's components. We have developed a set of phishing feature extraction criteria that we have applied to the process of getting new features. Classifying these characteristics was done using SVM, NB, and ELM. It was found that the ELM performed best when six distinct activation functions were tested.

REFERENCES

1. Anderson, R. (2020). Security Engineering: A Guide to Building Dependable Distributed Systems (3rd ed.). Wiley.
2. Mirkovic, J., Reiher, P., & Thomas, R. (2005). Internet Denial of Service: Attack and Defense Mechanisms. Prentice Hall.
3. Buczak, A. L., & Guven, E. (2016). "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," IEEE Communications Surveys & Tutorials, 18(2), 1153-1176.
4. Chandola, V., Banerjee, A., & Kumar, V. (2009). "Anomaly Detection: A Survey," ACM Computing Surveys (CSUR), 41(3), 15.
5. Sommestad, T., Ekstedt, M., & Johnson, P. (2010). "A Probabilistic Relational Model for Security Risk Analysis," Computers & Security, 29(6), 659-679.
6. Modi, C., Patel, D., Borisaniya, B., Patel, H., & Rajarajan, M. (2013). "A Survey of Intrusion Detection Techniques in Cloud," Journal of Network and Computer Applications, 36(1), 42-57.

INNO SPACE
SJIF Scientific Journal Impact Factor

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com