International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 10 | Oct - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Phishing Detection Using Machine Learning

Phishing Detection Using Machine Learning: An Intelligent Approach to Identifying Malicious URLs

Mrs. Diksha Rajendra Tanpure, Mr. Swaraj Sanjay Malve

Assistant Professor, Department of Master of Computer Application, JCEI's Jaihind Institute of Management and Research, Vadgaon Sahani Email: dikshatanpure33@gmail.com
swarajmalve79@gmai.com

Abstract - Phishing is a cyberattack in which attackers create fake websites, emails or messages to trick users into sharing sensitive information, such, as passwords, banking details and personal data. Conventional security methods may struggle to detect newly created or modified phishing attacks.Therefore machine learning can be used to improve the detection of phishing activities.This project, **Phishing Detection Using Machine Learning** aims to develop a system that identifies and classifies phishing and websites by analyzing their features. The system uses machine learning techniques to identify fraudulent websites by examining URL patterns, special characters, domain-related information and website security features. The available data is. A model was developed to identify and differentiate phishing websites from legitimate ones.The effectiveness of the model is assessed using performance measures, including accuracy, precision, recall and F1-score.The proposed system aims to provide more reliable phishing detection reduce the risk of online fraud and improve users awareness of cybersecurity threats. This project highlights the role of machine learning in improving online security by detecting suspicious websites and helping users avoid exposing sensitive information to potential cyber threats.

Keywords: Phishing Website Detection, Machine Learning, Cybersecurity, URL Feature Analysis, Classification Techniques, Web Security.

1. INTRODUCTION

Today the internet is used by people for online banking, shopping, learning, social networking and many other services.. The more we use the internet the more we face cyber threats. One of those threats is phishing. Phishing is a danger. In phishing bad people use websites or tricking messages to steal private information from users. They want login names, passwords, money details and personal data.Phishing detection using machine learning is a project that wants to find and label phishing websites as bad or safe. Machine learning lets a computer learn from website data and then guess if a new website is dangerous. The system looks at website details. It checks the URL shape the domain name, any symbols whether HTTPS is used and other useful clues. I notice that the system looks at website details.

The proposed system applies machine learning algorithms to analyze website features and identify potential phishing threats. This approach aims to improve phishing detection, reduce the risk of cyberattacks, and protect users from fraudulent websites and online scams.

Machine learning-based phishing detection reduces the need for manual website verification and enables faster identification of potentially malicious URLs. The primary objective of this project is to develop an efficient system that automatically detects phishing websites and improves online security.

to build an trustworthy phishing detection system. This system helps users see websites before they give away private data. I feel that this project helps us to see sites before giving away private data. This project highlights the application of machine learning in cybersecurity to enhance online safety and provide users with a more secure browsing experience.

1.1 Problem Statement

Phishing attacks are a threat to cybersecurity. Attackers set up websites that look just like real ones. Their goal is to take users' data, login passwords and money details. Old ways of spotting phishing sites often miss new. Changed scam sites.The main problem is to build a phishing detection system with machine learning. The system examines various website characteristics, including URL structure, domain information, unusual characters, and security indicators, to identify potential phishing threats.The goal is to decide if a site is phishing or real using patterns that the machine learns from data.The new system will make phishing detection more accurate and reliable. It aims to minimize the risk of online fraud and protect users from potential cyber threats.Give users safer protection when they surf the web.

1.2 Objectives

1. The goal is to create a system that uses machine learning to find phishing websites and tell them apart from websites.
2. The aim is to look at features of websites like the way the URL is written the domain details, strange symbols and security parts to spot websites that could be phishing.
3. The plan is to use the machine learning methods to sort websites into phishing or real.
4. The purpose is to make the system better at finding phishing websites in an accurate and trustworthy way.
5. The objective is to lower the chances of fraud and keep peoples personal information login details and money safe.
6. The objective is to develop an automated system for detecting malicious websites and strengthening overall cybersecurity.
7. The task is to check and compare machine learning methods with measures, like accuracy, precision, recall and F1-score to see how well they find phishing websites.

1.3 Contributions

This paper makes four main contributions: (1) a unified, modular reference architecture for AI-driven data quality monitoring that does not depend on any particular data source; (2) a comparative, empirical evaluation of statistical, machine learning, and deep learning anomaly detectors on structured data quality tasks; (3) a hybrid ensemble detection strategy that beats the individual models it draws from; and (4) a discussion of practical deployment concerns, covering alert design, explainability, and integration with existing data pipelines.

2. LITERATURE REVIEW

Phishing is a cybersecurity problem. Phishing attackers build websites to steal usernames, passwords and banking details. I have seen many researchers try ways to spot phishing websites and to make online security stronger.
Machine learning is an effective approach for analyzing website characteristics and identifying patterns associated with phishing activities.

Several studies have looked at algorithms such as Decision Tree, Random Forest, Support Vector Machine, Logistic Regression and Naive Bayes for phishing detection. These algorithms judge websites by checking features, like URL length, domain age, special characters whether HTTPS is used and other website attributes. Using these features can improve the accuracy and reliability of phishing website detection.

Researchers have also studied ways to select the features and to clean data before training. This helps detection work better. Saves computer time. Ensemble learning, where several algorithms work together is another approach to get classification. Phishing detection performance is usually measured by accuracy, precision, recall and the F1-score.
These numbers tell how well the system works.

With all this progress phishing tricks keep changing. This makes it hard for detection systems to spot every threat. Some models work well on data but may fail against new phishing websites that have not been seen before. That is why more research is needed to build machine learning systems that stay accurate fast and can adapt to phishing ways.This project will look at website features apply the machine learning algorithms and compare how well each one works to find phishing websites more effectively.

Research gap

Many machine learning methods have been made to spot phishing websites. Some problems still exist. Many current ways to find phishing websites can miss bad sites because bad hackers often change the look of their pages the URL patterns and the domain details to get past safety systems. Some models also rely on data sets, which can stop them from spotting new phishing tricks.Another gap in research is finding the website clues while still keeping detection strong and cutting down on computer work. Different machine learning algorithms can give results depending on the data set and chosen clues. So looking at machine learning algorithms in the same test setting is key.This project plans to fix those gaps by looking at website clues and testing good machine learning algorithms with scores like accuracy, precision, recall and F1. The goal is to find a way to separate phishing websites from real sites and boost trust, in detection.

3. METHODOLOGY

3.1 Data Collection

The data set comes from data repositories. It contains both genuine web URLs. These examples are used to train and test the machine learning model. The data set gives the machine learning model the needed examples to tell phishing URLs from legitimate URLs.

3.2 Data Preprocessing

The data set from the step is cleaned and ready for the next step. It is also adjusted so that there are no repeated rows, no inconsistencies and no missing data.

3.3 Feature Extraction

We pull out the features from the URLs in the data set to spot phishing URLs. The features we can use are URL length the number of dots in the URL the number of characters whether an IP address is used, HTTPS status and whether suspicious words appear.

3.4 Feature Selection

We choose the most relevant features from the extracted set. This improves the performance of the phishing detection machine learning model. Cuts down on the amount of information needed.

3.5 Model Training

The cleaned data set is split into training and testing sets. We then train machine learning models, like Decision Tree, Random Forest, Support Vector Machine and Logistic Regression to detect phishing URLs.

4. RESULTS

The proposed phishing detection system can analyze the website URL based on a machine learning process to detect both URLs and phishing URLs. This phishing detection system classifies URLs based on their features. Uses the same data to train the model to detect similar patterns.
Thus the system will be able to perform the following results:

1. Phishing Website Detection: The phishing detection system will be able to detect phishing websites by analyzing URLs.
2. Legitimate Website Detection: The phishing detection system will also be able to detect websites.
3. Model Evaluation: The performance of the machine learning model will be evaluated using accuracy, precision, recall and F1-score.
4. Improved Online Safety: The phishing detection system will help users identify websites and minimize phishing attacks.

5. DISCUSSION

A phishing detection system that uses machine learning technology can quickly tell the difference between phishing URLs and real URLs. The model is trained using features of URLs. It looks at patterns that appear during the training to decide if a website is safe or not.

The important part of building this kind of model is gathering a large amount of data for training. Choosing the features and the best algorithm also matters a lot. The performance of the model can be checked using accuracy, precision, recall and F1 score. Also cleaning the data and updating the model can help make the system better.

Overall this method can make browsing the web safer by making it easier to find phishing sites. However there are some problems with this kind of model like the chance of getting negatives and false positives. So keeping the data up, to date and improving the model are very important for the system to work well.

6. CONCLUSION AND FUTURE WORK

Phishing detection using machine learning can find websites that are trying to trick users. It helps tell the difference between phishing websites and real websites. This method looks at the features of URLs to decide if they are safe or dangerous. The performance of machine learning models is checked using accuracy, precision, recall and F1-score. Using machine learning, for phishing detection can make the internet safer by stopping phishing scams.

Future studies can use machine learning to check URLs in time. This would allow phishing detection to work quickly as users browse. It is possible to add phishing detection tools into web browsers. These tools can warn users when they are visiting a website. More research can help make phishing detection faster. Better methods will lead to reliable detection and stronger online security.

REFERENCES

1. Sahingoz, O. K., et al. (2019). Machine learning based phishing detection from URLs.

2. Jain, A. K., & Gupta, B. B. (2018). Towards detection of phishing websites on client-side using machine learning based approach

3. Aljofey, A., et al. (2022). An effective detection approach for phishing websites using URL and HTML features.

4. A stacking model using URL and HTML features for phishing webpage detection. (2019).

5. Phishing Websites Detection using Machine Learning with URL Analysis. (2022).

6. A high-accuracy phishing website detection method based on machine learning. (2023).

7. Safi, A., & Singh, S. (2023). A systematic literature review on phishing website detection techniques.

8. Phishing Website Detection With Semantic Features Based on Machine Learning Classifiers. (2022).

9. Modeling Hybrid Feature-Based Phishing Websites Detection Using Machine Learning Techniques. (2022

10. An In-Depth Benchmarking and Evaluation of Phishing Detection Research for Security Needs. (2020).

11. Phishing or Not Phishing? A Survey on the Detection of Phishing Websites. (2023).

12. Mohammad, R., & McCluskey, L. (2012). Phishing Websites Dataset. UCI Machine Learning Repository.

13. Feature Engineering for Phishing Website Detection Using Machine Learning: A Systematic Review. (2025). IEEE Access.

14. Phishing URL Detection in Real-Time: Machine Learning and Deep Learning Approaches on a Custom Threat Intelligence Dataset. (2026). Procedia Computer Science.

15. Phishing Website Detection Using Machine Learning. IEEE Xplore.