



Dr. Vishwanath Karad
**MIT WORLD PEACE
UNIVERSITY** | PUNE
TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS



TEAM SYNTAX TERMINATORS

School of Computer Science and Engineering

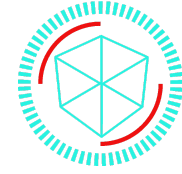


Team Members :-
Kanishk Jamgaonkar
Nidhi Sabu
Muhammad Arab

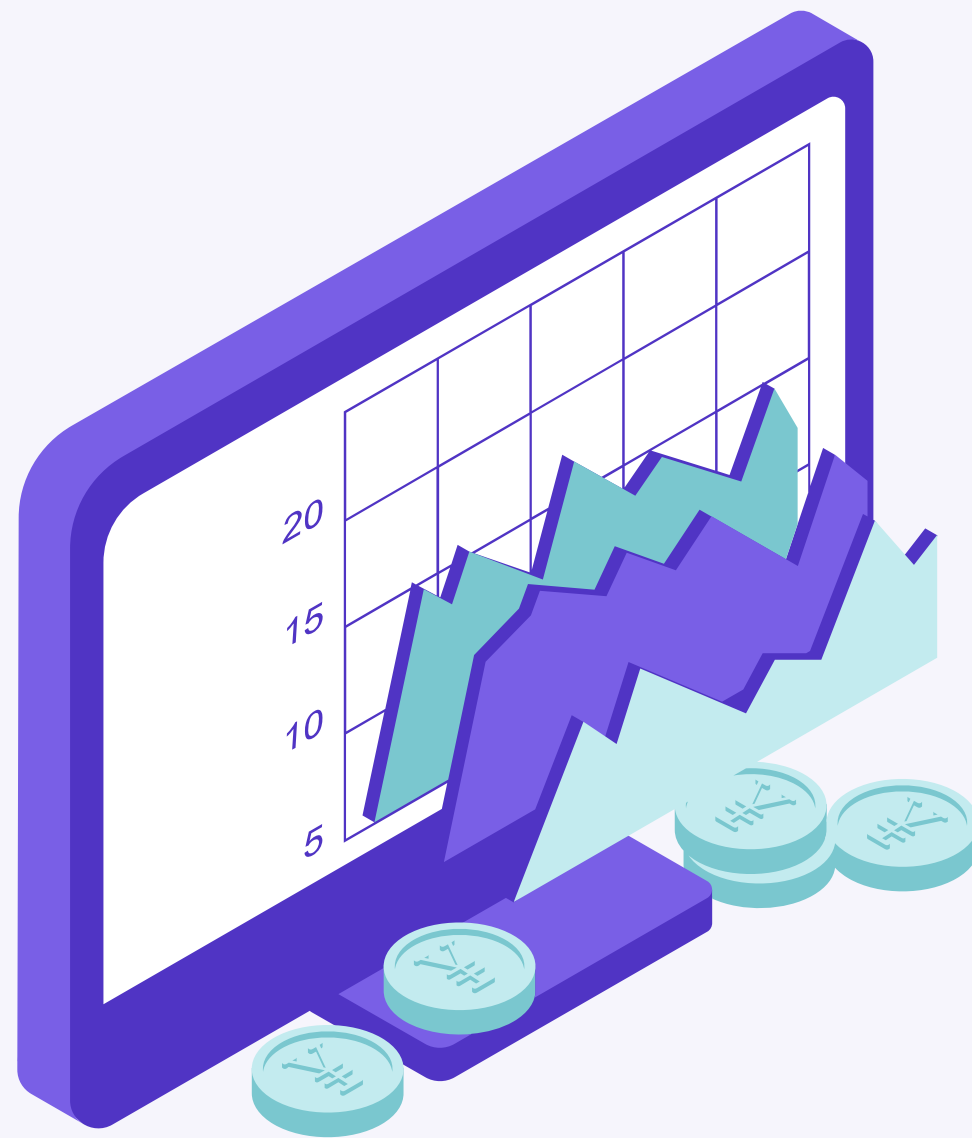


SecureSite

Machine learning powered, real-time phishing
and fraudulent website detection.

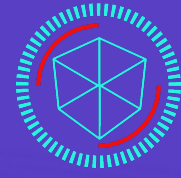


Overview



In this digital age, protecting online users from becoming victims of online fraud, divulging confidential information to an attacker is very important.

Phishing is one of the most serious cyber-attacks. It must be identified as early as possible to minimise the damage.



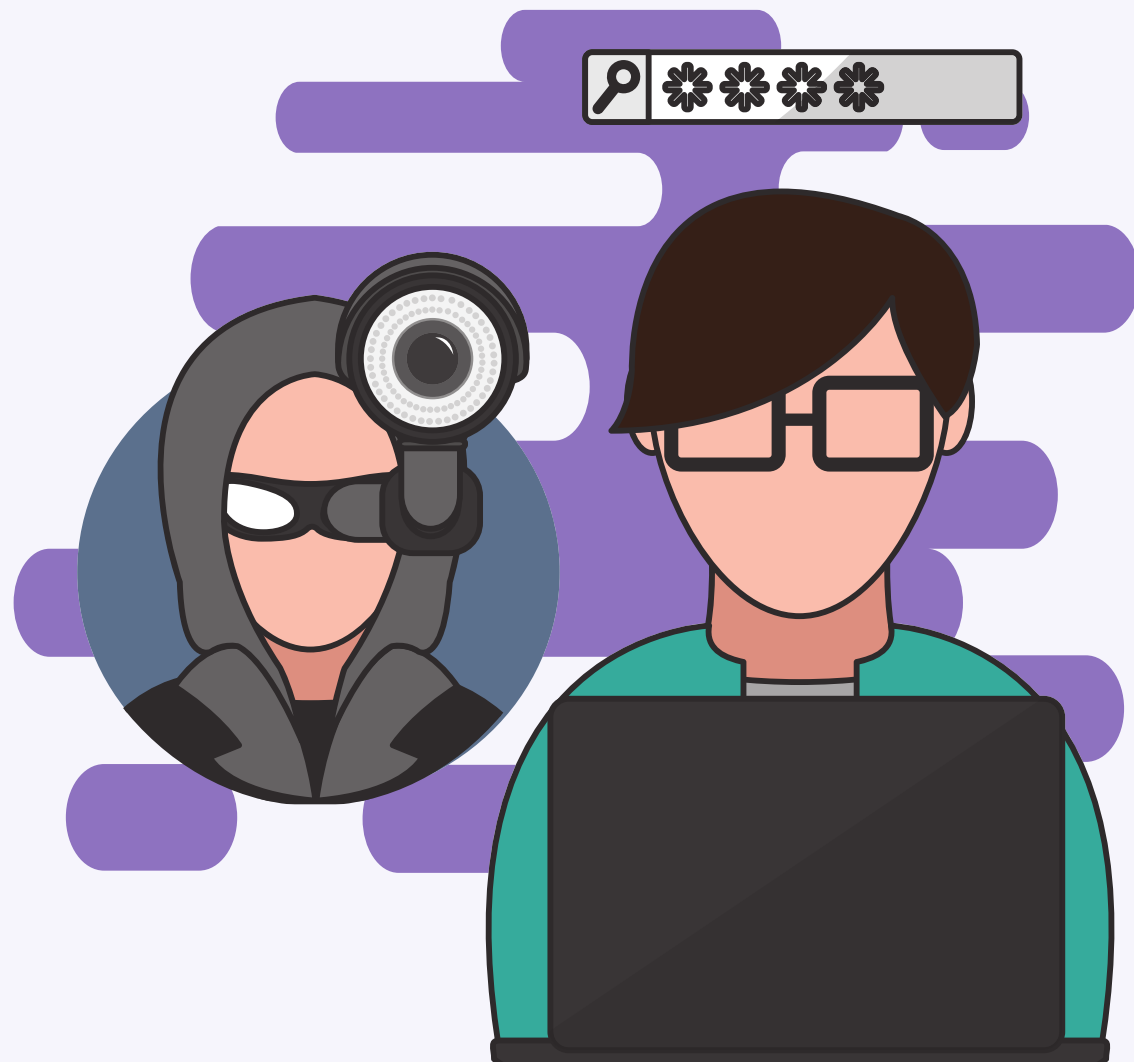
Problem Statement

Phishing is one of the techniques which are used by intruders to get access to the user credentials or to gain access to sensitive data. This type of access is done by creating a replica of the websites which looks the same as the original websites which we use on a daily basis but when a user clicks on the link he will see the website and think it's original and try to provide his credentials.

To overcome this problem we are using some of the machine learning algorithms in which will help us to identify the phishing websites based on the features present in the algorithm. By using this algorithm we can be able to keep the user personal credentials or sensitive data safe from intruders.



Objective



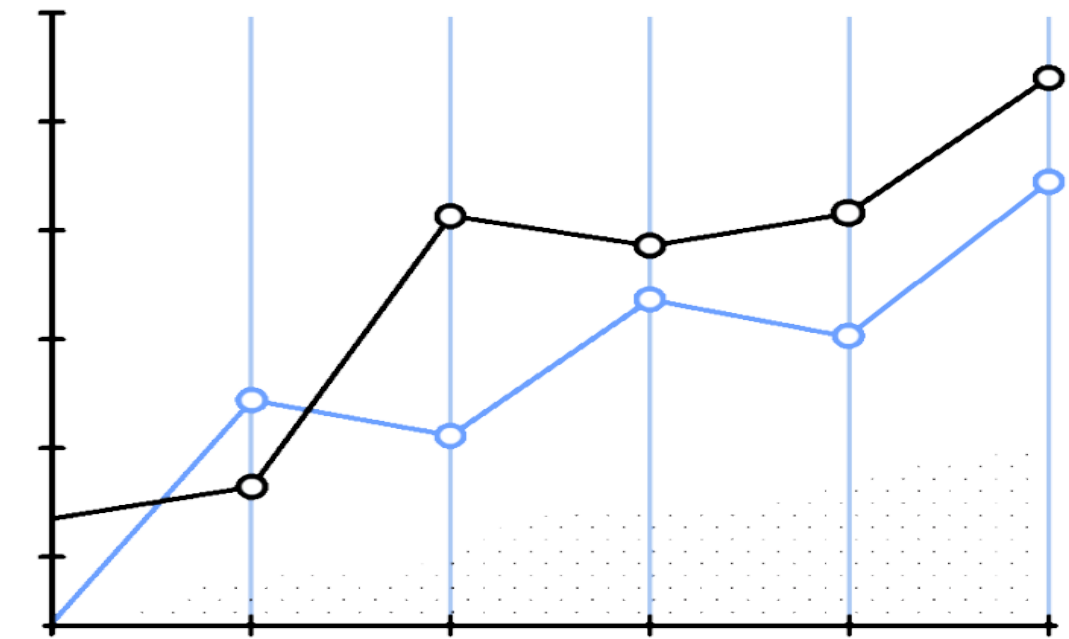
The main objective of this project is to build a model which analyses the dataset which is able to identify the phishing URLs with respect to their attributes.

We will extract the features from the given URL and give them to our model for predictions. We are going to compare different machine learning algorithms and the best classifier that will be used to detect malicious URL .



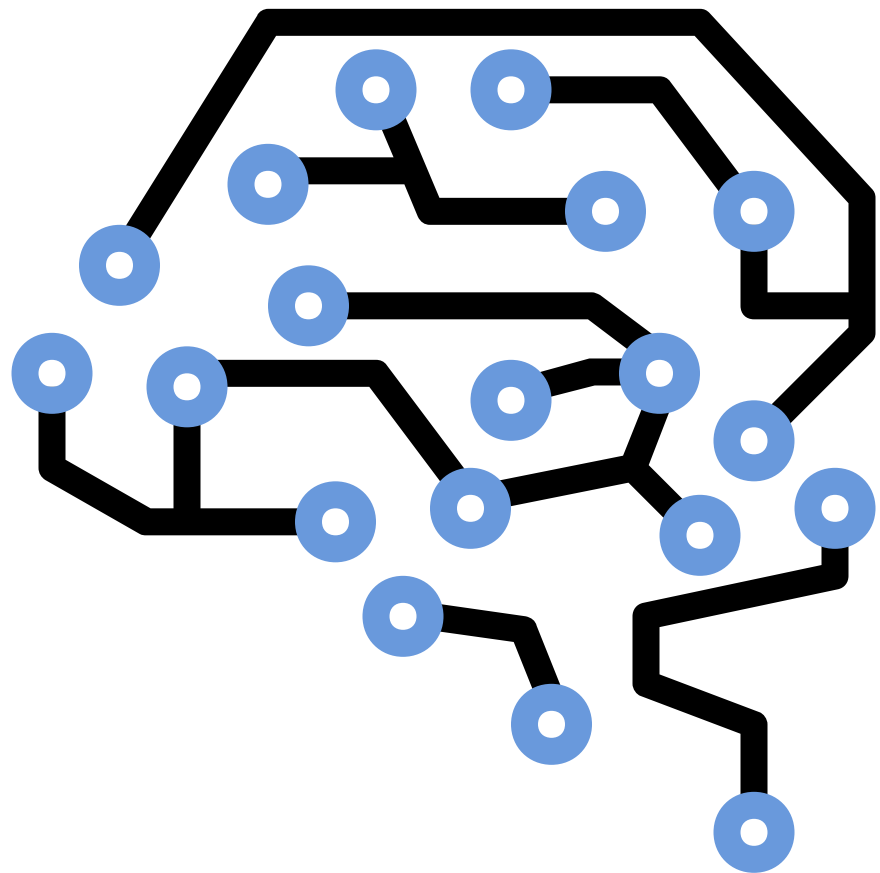
Dataset

To evaluate our machine learning techniques, we have used the 'Phishing Websites Dataset' from the UCI Machine learning repository. It consists of features of 11055 URLs. Each URL has 30 Features/Attributes .



Data Set Characteristics:	N/A	Number of Instances:	2456	Area:	Computer Security
Attribute Characteristics:	Integer	Number of Attributes:	30	Date Donated	2015-03-26
Associated Tasks:	Classification	Missing Values?	N/A	Number of Web Hits:	190571

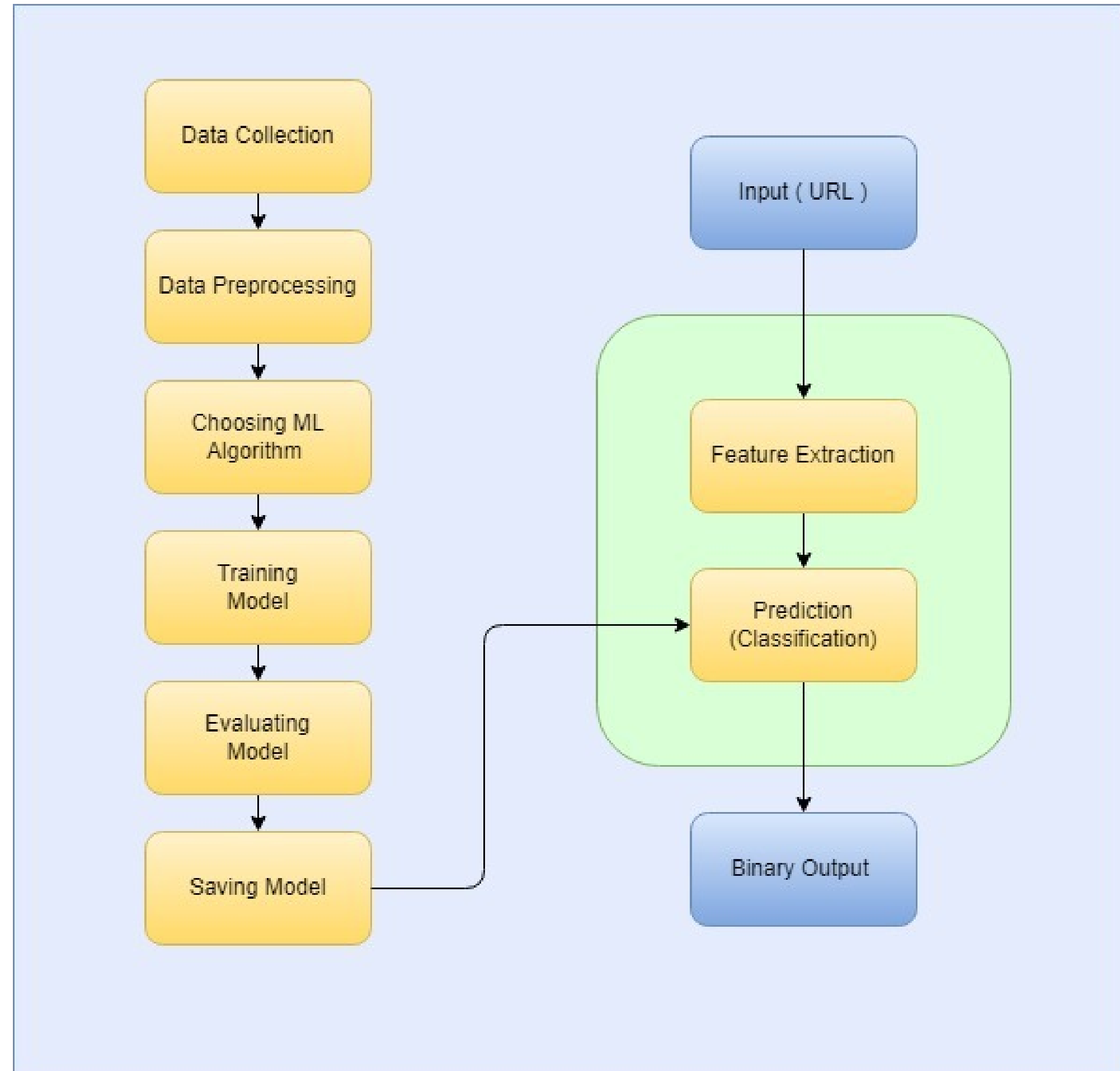
Algorithms Used



- 01 Logistic regression
- 02 K-Nearest Neighbors Algorithm
- 03 Naive Bayes Classifier
- 04 Random Forest Classifier
- 05 Artificial Neural Network

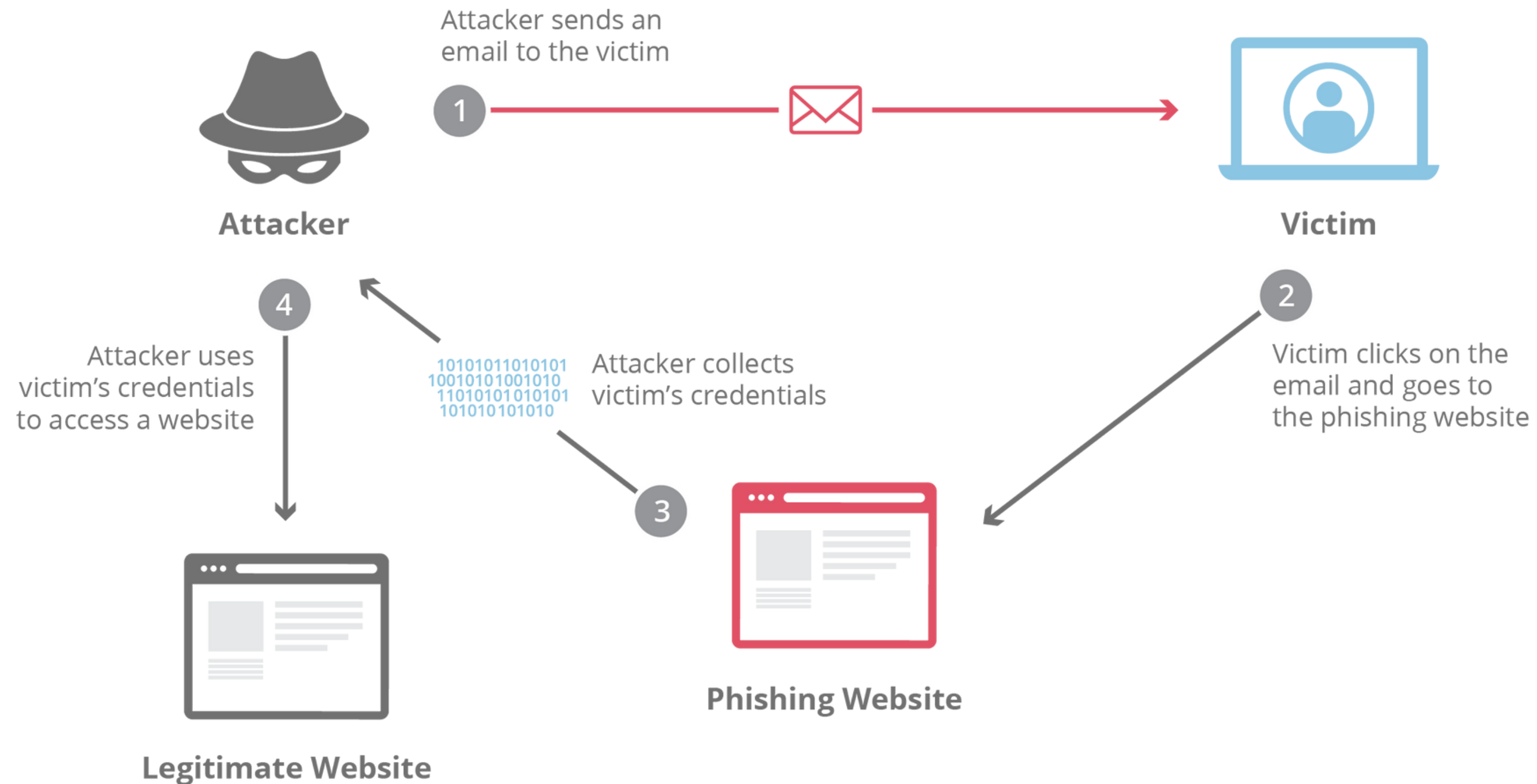


System Architecture





How does Phishing work?



Implementation (Outputs)

hackmitwpu.com

Check



This site is secure

Go back

2020logs.duckdns.org

Check

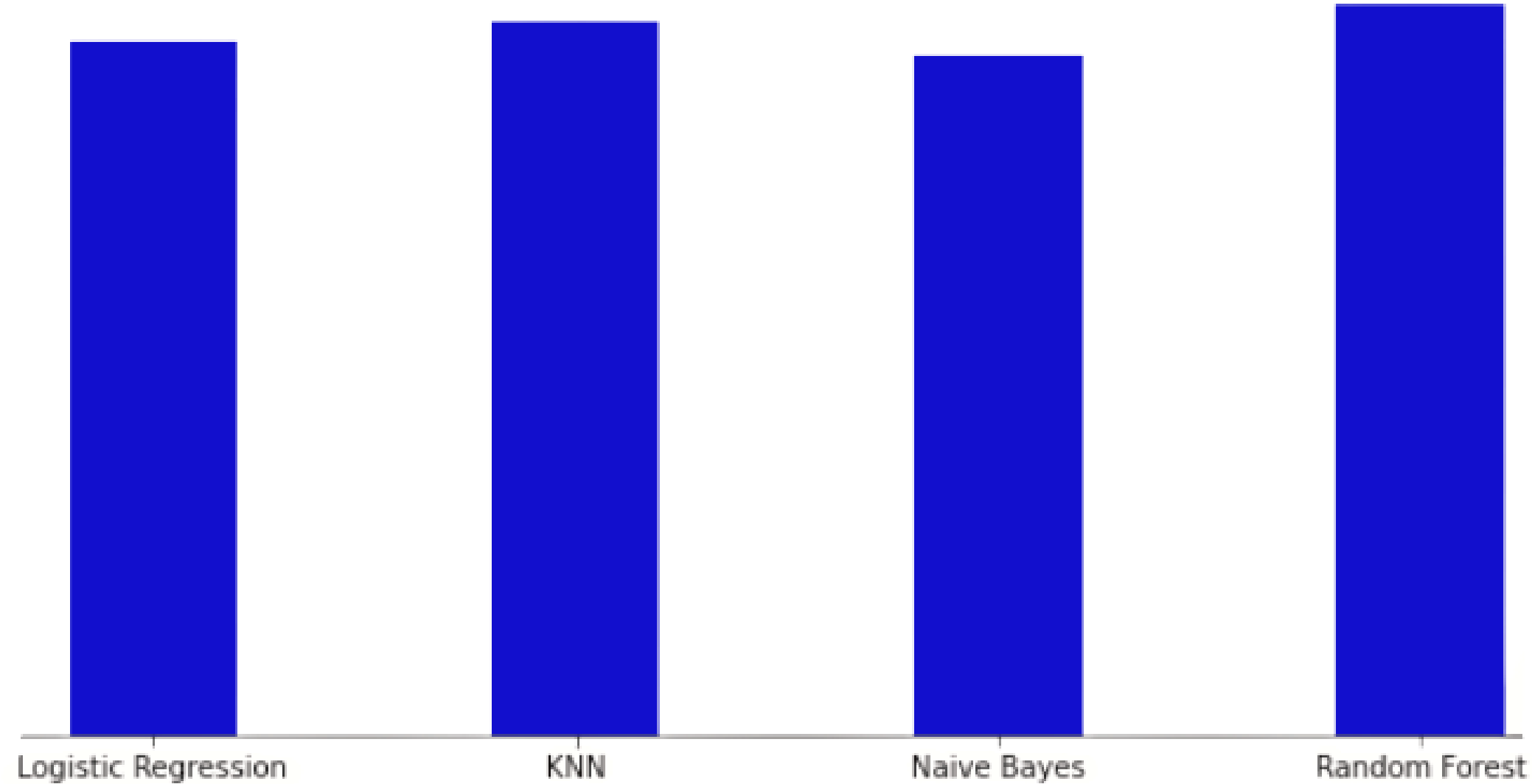


This site is **not** secure

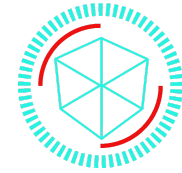
Go back



Visualization Screenshots



- 01 Logistic regression = 0.92655
- 02 K-NN Algorithm = 0.95115
- 03 Naive Bayes Classifier = 0.90484
- 04 Random Forest Classifier = 0.97395

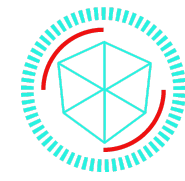


Impact

Phishing is a type of attack that uses fake websites and email campaigns to deceive people into giving up personal information, downloading malware/ransomware, or both. Phishing attacks are also carried out on a frequent basis against organizations.

Many businesses rely too much on basic spam filtering software to prevent such attacks; nonetheless, the majority of spam filters will eventually fail against a well-designed phishing site.

This URL-phishing-detector will help identify phishing sites, without posing any risk to the consumer or the organization.



Conclusion

Thus to summarize, we have seen how phishing is a huge threat to the security and safety of the web and how phishing detection is an important problem domain.

We have reviewed some of the traditional approaches to phishing detection; namely blacklist and heuristic evaluation methods, and their drawbacks.

We have tested four machine learning algorithms on the 'Phishing Websites Dataset' from the UCI Machine Learning Repository and reviewed their results.

We then selected the best algorithm based on its performance and built a Django Webapp for detecting phishing web pages.

The Webapp allows easy deployment of our phishing detection model to end-users.

Code Link

Github

<https://github.com/kanishk7559/hackmitwpu>

Thank you !