Detecting Malicious and Benign Websites

Proposal issued:
04.11.2021

Prepared by:

HADEEL AMIN AL REHAILI

P - 0555758522 | **E –** alrehailiha1.ngha.med.sa – hadeel.03a@gmail.com

OVERVIEW OF PROJECT

One of the biggest challenging on the internet now is web Security; there are billions of websites active on Internet, So threats on the Internet is increasing and the hackers evolving newer techniques to trap web users. With the advancement of technology, especially in the field of AI and ML, we can detect malicious websites by using machine learning.

Therefore, the **Goal** of the project is to build a ML model that can detect these attacks based on previous attacks happened.

The **Benefits** of building such this model is to help organization in improving their security on premises and providing a feedback to their security tools to be able to recognize the malicious attack.

The data was gathered from Web between November 2019 and March 2020, and it is hosted in public repository with Direct URL to data, Link to Data.

Data Description

The dataset I will work with is a data that contain information about Malicious and benign Websites. The data contains 5000 points with nine features. I expect to work with categorical and numerical types of feature. Every website has features like, IP Address, URL length, geographical location of the IP, protocol used, Domain of the website and many other features that detect where the website is safe or not.

Algorithms

I am planning to choose the Logistic regression Algorithms as a classification method to predict the target of the model. The target of my model is whether the website is “Good” (1) or “Bad” (0).

Tools

The tools I expect to use in this project would be as following:

- **Jupyter notebook as IDE.**
- **Pandas and numpy to data processing and cleaning.**
- **Matplotlib and Seaborn for plotting.**
- **Scikit-learn for modeling.**