Hewlett Packard
Enterprise

MACHINE LEARNING HACKATHON CG 2022

Team Name- Silent_Coders

Team Leader Name- Ashish Kumar

Team Leader Email Address- ashishok88@gmail.com

BRIEF DESCRIPTION OF THE PROBLEM AT HAND:

- As each coin has two sides, so does technology. On one hand, with the advancement in technology the lives are getting better, on the other hand, the ill use of technology is also increasing. The suspicious activities are increasing ranging from dos attacks, phishing, hacking etc. In this hackathon, we are concerned with the increasing phishing attacks.
- Typically, Phishing is a type of social engineering where an attacker sends a fraudulent (e.g., spoofed, fake, or otherwise deceptive) message designed to trick a person into revealing sensitive information to the attacker or to deploy malicious software on the victim's infrastructure like ransomware.

Solution proposed and description:

The main task is to overcome this Phishing attack at different sites and organization .

So I have created a model based on the data analysis to identify if the website is legitimate or a phishing website.

Technology/Tool Stack Used:

- VS code editor
- Python
- LogisticRegression() model to predict the result

Approach

- The Problem statement is very clear i.e. to identify the Phishing attacks over the websites. So first thing is to describe the dataset with its different parameters and characteristic of the dataset, then I will use train, test split method to separate the training data and test data. I will be using the LogisticRegression() model in order to predict the result. As '1' represents the legitimate , '0' represents suspicious, '-1' represents phishing. Accuracy score for the Training data is 91.9% and for the Testing data is 91.5% which is very good as it will predict the best result. Then the final step is to separate the 'Result' and 'Key' column to a CSV file for training data and Testing Data.

SOURCE CODE AS ZIP OR GITHUB URL:

https://github.com/ashishok/Techgig_ML

THANK YOU