ABSTRACT - MAIN PROJECT

RIYA VINCENT
ROLL NO: MCA431

PHISHING ATTACK DETECTION USING MACHINE LEARNING

In recent years web applications that are hacked every day are estimated to be 30 000, and in most cases, web developers or website owners do not even have enough knowledge about what is happening on their sites. Web hackers can use many attacks to gain entry or compromise legitimate web applications, they can also deceive people by using phishing sites to collect their sensitive and private information. In response to this, the need is raised to take proper measures to understand the risks and be aware of the vulnerabilities that may affect the website and hence the normal business flow. Mitigations against the most common web application attacks are set, and the web administrator is provided with ways to detect phishing links which is a social engineering attack, the study also demonstrates the generation of web application logs that simplifies the process of analyzing the actions of abnormal users to show when behavior is out of bounds, out of scope, or against the rules.

The methods of mitigation are accomplished by secure coding techniques and the methods for phishing link detection are performed by various machine learning algorithms and deep learning techniques. This application will be tested and evaluated against various attack scenarios, the outcomes obtained from the test process showed that the website had successfully mitigated these dangerous web application attacks, and for the detection of phishing links part, a comparison is made between different algorithms to find the best one, and the outcome of the best model gave 98% accuracy.

Collection of the data for testing purposes. We will collect the data from the dataset and then use it for classification.Then splitting the data into train and test sets. These sets will be divided in the ratio of 80:20. The train set will be used in training the model and the test set will be used in verifying our model.Building the model is applying the algorithms, i.e., implementing the ANN and the XGBoost algorithms.Training the model using the test set which will train the ANN and the XGBoost models.Importing links to the prediction step is the main step after the training of the model. This is testing the model for accuracy and checking whether the model is implemented successfully or not.Deploying the model so that the model can be used for the checking of phishing websites in the real world.