**ASSIGNMENT 1.**

**NAME**: STANLEY NGUGI CHEGE **REG NO**: SCT212-0065/2017

**Task**:

*Statement of the problem; Why choice of topic and problem; Purpose statement and objectives; How you propose to solve the problem; Expected project outcome and their potential.*

# Statement of the problem;

Internet fraudsters, an elite team of cyber-criminals, are individuals who make use of technology to commit malicious activities on the internet with the intention of personal gain such as making profits. The absence of laws that restrict the acquisition of domains for online businesses and the ease of acquiring them paves way for internet fraudsters who acquire domains with malicious intentions. They are able to disguise their identities and IP addresses through the use of Virtual Private Networks and proxies which makes it difficult to track them once their deception comes to light. Most internet users are not computer security specialists and are therefore vulnerable to an attack. A phishing email would easily capture their attention and they will fall under the control of the phisher. The problem can be minimized by addressing it in two folds; developing more targeted anti-phishing detection and interventions techniques; implementing a round-the clock protection mechanism to notify internet users in case they come across fraudulent URLs.

# Why choice of topic and problem;

As Internet fraudsters evolve and sharpen their skills, it becomes more difficult for novice users to detect or distinguish phishing websites from legitimate ones. Therefore, there is need to provide anti-phishing solutions that adopt Machine Learning which is more practical and effective in combating phishing. Machine Learning anti-phishing techniques rely on website features to derive knowledge that can assist in identifying phishing websites. Phishing campaigns usually take significantly lower times in their attack (a few hours) such that relying on blacklists and whitelists does not guarantee their detection. Therefore, a binary classification of websites as either legitimate or fraudulent in real time will help internet users stay ahead of the game as they protect themselves from cyber attacks

# Purpose statement and objectives;

The main objective of the study is to make use and evaluate the efficiency of Supervised Machine Learning in the detection and classification of websites as fraudulent or legitimate. To achieve this, the following objectives will be met:

* To train a machine learning model using a dataset to classify websites as legitimate or fraudulent
* To collect, analyze and organize training and testing dataset.
* To identify common distinguishing features between fraudulent and legitimate websites
* To provide a round-the-clock protection mechanism to keep the user aware of suspicious websites

# How you propose to solve the problem;

Once all the required data has been gathered, cleaning, transforming, and modeling operations will be conducted on the data to discover useful information. A script to extract various distinguishing features of a URL such as the domain name (or IP address), protocol and the suffix will be written. The dataset will be analyzed using Exploratory Data Analysis tools. Data will be divided into training and testing sets. Training a machine learning model to distinguish phishing URLs from legitimate ones will be conducted. The model will then be tested using test data before deployment. A browser extension will be developed to ease with user interaction.

# Expected project outcome and their potential.

The problem can be minimized by addressing it in two folds; developing more targeted anti-phishing detection and interventions techniques using a trained classification machine learning model; implementing a round-the clock protection mechanism to notify internet users in case they come across fraudulent URLs through development of a browser extension. This approach at combating phishing is more practical and will give internet users more confidence while transacting with online platforms.