ABSTRACT

Phishing, malware, begin, defacement attack is a simplest way to obtain sensitive information from innocent users. With the rise of malicious activities on the internet, it has become essential to have intelligent systems in place that can detect and identify such activities. URL analysis has been a proven method for identifying attacks such as phishing and malware. In the past, various approaches have been used that include lexical features, network traffic, and hosting information to categorize URLs. Aim of the phishers is to acquire critical information like username, password and bank account details. This paper deals with machine learning technology for detection of phishing URLs by extracting and analyzing various features of legitimate and phishing URLs. XGB, Random Forest and LGB classifier are used to detect phishing websites. Aim of the paper is to detect phishing URLs as well as narrow down to best machine learning algorithm by comparing accuracy rate, false positive and false negative rate of each algorithm, However, these approaches are time-consuming and can cause significant delays in real-time systems explore a lightweight approach to classify malicious websites using only lexical URL analysis. The objective is to determine the upper limit of classification accuracy that can be achieved through purely lexical methods.