Phishing Website Detection Using Support Vector Machines and Nature-Inspired Optimization Algorithms

Anupam, Sagnik (School: Delhi Public School, R. K. Puram)

Phishing websites are amongst the biggest threats Internet users face today, with an average of over 1.3 million such websites being created every month, according to 2017 Webroot Data. Existing methods like blacklisting, using SSL certificates, etc. often fail to keep up with the increasing number of threats. This project aims to utilise different properties of a website URL, and use a machine learning model to classify websites as phishing and non-phishing. These properties include the IP address length, the authenticity of the HTTPs request being sent by the website, usage of pop-up windows to enter data, Server Form Handler status, etc. A Support Vector Machine binary classifier trained on an existing dataset was used to predict if a website was a legitimate website or not, by finding an optimum hyperplane, which separated the two categories when the features were plotted on a graph. This optimum hyperplane was found with the help of four optimization algorithms, the Bat Algorithm, the Firefly Algorithm, the Grey Wolf Algorithm and the Whale Optimization Algorithm, which are inspired by various natural phenomena. My research shows promising results, with the four algorithms displaying 90.18% mean classification accuracy and a mean F1 score of 0.88. Amongst the four algorithms, it was determined that the Grey Wolf Algorithm's performance was statistically better than that of the Firefly Algorithm, but there was no statistical difference while comparing the performance of any other pair of algorithms.