

Phishing Site Detection Plugin

Using RandomForest classifier in Javascript

Dataset: <https://archive.ics.uci.edu/ml/datasets/phishing+websites>¹

Phishing is defined as mimicking a creditable company's website aiming to take private information of a user. Based on the IEEE paper, *Intelligent phishing website detection using random forest classifier*², the random forest classifier seems to outperform other techniques in detecting phishing sites.

These techniques, however use python machine learning libraries and thus they can't be used in the browser in real-time. The main objective of this project is to develop a Chrome browser plugin that detects phishing sites in real-time while the user browses the page. One common approach is to make the prediction in a server and then let the plugin to contact the server for each page. Unlike the old approach, this project aims to run the classification in the browser itself. The Advantage of classifying in the client side browser has advantages like, better privacy (the user's browsing data need not leave his machine), independent of network latency.

This project is mainly of implementing the above mentioned paper in Javascript for it to run as a browser plugin. Since javascript doesn't have much ML libraries support and considering the processing power the client machines, the approach needs to be made lightweight.

¹ <https://archive.ics.uci.edu/ml/datasets/phishing+websites>

² <https://ieeexplore.ieee.org/abstract/document/8252051/>