Grandpa Phishing Shield

# Classification Problem, 100pts

A good method to steal your information is to deceive you into a fake web page that looks as a well-known page. The naiveness of people, especially those older or not tech-savvy, will be easily fooled into these pages and their information stolen.

As a data scientist, you have the challenge of identifying these spam web pages so you can alert those browsing into them.

OBJECTIVE:

Define an algorithm to correctly classify a web page as being spam or not.

FEATURES:

For each web page 139 features were calculated.

TRAIN SET:

The train set has 6169 pre-classified web pages.

TEST SET:

The test set consists of 475 not classified web pages.

LINK TO DATA:

DELIVERABLES:

* Summary (must): Explain your approach to solve the problem to any hackathon supporter of the Data Science challenge.
* Classification: Share your classification results of the test set. Include only the ID and CLASS columns. You will get back your F1 score\*

**\*** Your F1 score represents the number of points you will obtain. Zero being 0 points and One being 100 points. You can share your classifications as many times as you wish, your highest value will be reported as the number of points obtained.