# Phishing Websites Dataset

**Dataset files**

* <http://archive.ics.uci.edu/ml/machine-learning-databases/00327/Training%20Dataset.arff>

**Dataset entry description**

Attribute 1 (having\_IP\_Address) : If an IP address is used as an alternative of the domain

name in the URL  
Attribute 2 (URL\_Length) : Use of long URL  
Attribute 3 (Shortining\_Service) : Use of URL shortening service   
Attribute 4 (having\_At\_Symbol) : Use of @  
Attribute 5 (double\_slash\_redirecting) : Use of //

Attribute 6 (Prefix\_Suffix) : Use of -

Attribute 7 (having\_Sub\_Domain) : Use of multiple sub domains i.e http://www.hud.ac.uk/students

Attribute 8 (SSLfinal\_State) : Use of HTTPS   
Attribute 9 (Domain\_registeration\_length) : Registered domain length

Attribute 10 (Favicon) : Use of graphic image  
Attribute 12 (port) : Usage of port 443 to bypass firewall and other blockers.  
Attribute 13 (HTTPS\_token) : Usage of HTTPS as a token i.e http://https-www-paypal-it-webapps-mpp-home.soft-hair.com/

Attribute 14 (Request\_URL) : Copy content from an original website i.e video, image etc..  
Attribute 15 (URL\_of\_Anchor) : If anchor tag using different domain url

Attribute 16 (Links\_in\_tags) : If links tag are using other domain links. i.e script, meta etc.

Attribute 17 (SFH) : After submitting a page we receive about:blank  
Attribute 18 (Submitting\_to\_email) : If informations are submitted to email  
Attribute 19 (Abnormal\_URL) : If the hostname is not include

Attribute 20 (Redirect) : If webpage redirecting to other webpage  
Attribute 21 (on\_mouseover) : If onMouseOver popup a fake url  
Attribute 22 (RightClick) : If onMouseOver for blocking right click.  
Attribute 23 (popUpWidnow) : If popup window asks for personal details

Attribute 24 (Iframe) : If iframe embed with an external webpage  
Attribute 25 (age\_of\_domain) : If age of domain < 6  
Attribute 26 (DNSRecord) : If DNS record not found

Attribute 27 (web\_traffic) : If web traffic is low and not found in Alexa rank in top 100000  
Attribute 28 (Page\_Rank) : If page is without rank

Attribute 29 (Google\_Index) :If not found in Google index  
Attribute 30 (Links\_pointing\_to\_page):If link is pointing to a another link  
Attribute 31 (Statistical\_report) :If webpage status reports are not good

Attribute 32 (Result) :

For more details : http://archive.ics.uci.edu/ml/machine-learning-databases/00327/Phishing%20Websites%20Features.docx

**Example entry**

-1,1,1,1,-1,-1,-1,-1,-1,1,1,-1,1,-1,1,-1,-1,-1,0,1,1,1,1,-1,-1,-1,-1,1,1,-1,-1

**Labelling field**

Attribute 32 (Result)

Possible values

* 1 means phishing website
* -1 means not phishing website

**Questions to answer using ML**

1. Is a website legitimate or phishing website?
2. How many phishing websites have @ sign in them?

**Reference**

* <http://archive.ics.uci.edu/ml/datasets/Phishing+Websites>
* <http://eprints.hud.ac.uk/16229/1/The_7th_ICITST_2012_Conference_-An_Assessment_of_Features_Related_to_Phishing_Websites_using_an_Automated_Technique.pdf>