

This project is constructed of 4 types of features

- (1) Domain-based features. Such as WHOIS information, domain and its sections, days after the domain was registered and remaining days before the domain expires.
- (2) Webpage-based features and JS properties. Such as disabled right click.
- (3) Lexical Features. Such as the number of dots in URL, number of parameters, TLD length.
- (4) HTTP Headers. Such as if a cookie is being set.

All 63 features used in this system are list below:

TABLE 1

The description of all 63 features

Type	feature	description
Domain	dns_record	DNS record in WHOIS database.
	check_age_of_domain	Days since the domain was registered in WHOIS database.
	check_ssl	Check for valid SSL certificate.

days_till_expiration_domain

Days till expiration of the SSL certificate

**Webpage
based**

favicon

If the favicon is loaded from another domain.

iFrame

IFrame tag.

rightClick

If the rightClick is disabled.

onmouseover

If the mouseOver property show different URL than the redirected.

check_URL_of_anchor

If a tags direct to different domain or nowhere.

sfh

No actions after submitting form.

a_tags_count

Number of a tags in the html of the webpage.

form_tags_count

Number of form tags in the html of the webpage.

email_tags_count	Number of email tags in the html of the webpage.
pass_tags_count	Number of password tags in the html of the webpage.
hidden_tags_count	Number of hidden tags in the html of the webpage.
actions_tags_count	Number of action tags in the html of the webpage.
signin_tags_count	Number of signin tags in the html of the webpage.
signup_tags_count	Number of signup tags in the html of the webpage.

Lexical

double_slash	More that 1 occurance of "/" in the URL.
qty_dot_url	Number of "." in the URL.
qty_hyphen_url	Number of "-" in the URL.

qty_questionmark_url	Number of "?" in the URL.
qty_at_url	Number of "@" in the URL.
qty_hashtag_url	Number of "#" in the URL.
qty_dollar_url	Number of "\$" in the URL.
qty_percent_url	Number of "%" in the URL.
TLD_length	Number of letters of the TLD(Top Level Domain)
TLD_count	Number of sub-TLDs
URL_length	Length of the whole URL
email_in_url	If there is an email in the URL.

word_script_in_url

If the word "script" is in the URL.

check_https_in_url

If "https" is in the URL.

qty_dot_domain

Number of "." in the Domain.

qty_hyphen_domain

Number of "-" in the Domain.

count_vowels

Number of vowels in the Domain

Domain_length

Length of the Domain.

Ip_in_domain

Ip in Domain.

Client_or_Server_in_domain

Words "client" or "server" in Domain.

qty_dot_directory

Number of "." in the Directory.

qty_hyphen_directory	Number of "-" in the Directory.
qty_dot_directory	Number of "." in the Directory.
qty_hyphen_directory	Number of "-" in the Directory.
qty_at_directory	Number of "@" in the Directory.
qty_slash_directory	Number of "/" in the Directory.
qty_percent_directory	Number of "%" in the Directory.
directory_length	Length of the Directory.
qty_dot_File	Number of "." in the File.
qty_hyphen_File	Number of "-" in the File.

qty_at_File	Number of "@" in the File.
qty_percent_File	Number of "%" in the Directory.
File_length	Length of the File.
qty_dot_params	Number of "." in the Parameters.
qty_hyphen_params	Number of "-" in the Parameters.
qty_at_params	Number of "@" in the Parameters.
qty_underline_params	Number of "_" in the Parameters.
qty_hashtag_params	Number of "#" in the Parameters.
qty_dollar_params	Number of "\$" in the Parameters.

qty_percent_params	Number of "%" in the Parameters.
--------------------	----------------------------------

params_length	Length of the Parameters.
---------------	---------------------------

tld_params	If there is a TLD in the Parameters
------------	-------------------------------------

count_params	How many parameters are being set.
--------------	------------------------------------

HTTP Headers

cookie	If a cookie is being set.
--------	---------------------------

strict_trans_sec	Existence of Strict_Trans_Sec header.
------------------	---------------------------------------

All features are numeric.