

3.1.2 Data collection. The uci repository provides secondary data used for the study. Data obtained by 2,455 phishing sites, which will be divided into two as training data as much as 80% or 1964 data and test data as much as 20% or 491 of data [11].

Data phishing site has 18 features [11]. Such features are as follows: having an IP address, URL length, shortening service, having @ symbol, the double slash redirecting, prefix suffix, having subdomains, SSL final state, domain registration length, favicon, port, HTTPS token, request a URL, the URL of anchor, links in tags, SFH, submitting to email and abnormal URL. This feature will be categorized as shown in Table 1.

Table 1. Features

Parameter	Features	Value Features
1 = valid 0 = suspicious -1 = phishing	Having IP Address	-1,1
	URL Length	1,0, -1
	Shortening Service	1, -1
	Having @ symbol	1, -1
	Double Slash Redirecting	-1,1
	Prefix Suffix	-1,1
	Having Sub Domains	-1,0,1
	SSL final State	-1,0,1
	Domain Registration Length	-1,1
	Favicon	1, -1
	Port	1, -1
	HTTPS Token	-1,1
	request URL	1, -1
	URL of Anchor	-1,0,1
	Links in Tags	1, -1,0
	SFH	-1,1,0
	Submitting to Email	-1,1
	abnormal URL	-1,1

Source: uci repository

From the data in Table 1, to determine phishing sites with non phishing sites using 18 features with the following explanation:

1. Having an internet protocol (IP) Address
If the IP address is used as the domain of the URL, such as <http://125.98.2.142/contoh.html>, it can be suspected that attempts to steal information. The IP address can also be converted into hexadecimal code.