









Dataset Characteristics

Multivariate

Associated Tasks

Classification

Instances

1353

Subject Area

Computer Science

Feature Type

Integer

Features

9

Dataset Information



The phishing problem is considered a vital issue in the e-commerce industry especially e-banking and e-commerce taking the number of online transactions involving payments. We have identified different features related to legitimate and phishy websites and collected 1353 different websites from difference sources. Phishing websites were collected from Phishtank data archive (www.phishtank.com), which is a free community site where users can submit, verify, track and share phishing data. The legitimate websites were collected from Yahoo and starting point directories using a web script developed in PHP. The PHP script was plugged with a browser and we collected 548 legitimate websites out of 1353 websites. There is 702 phishing URLs, and 103 suspicious URLs.

When a website is considered SUSPICIOUS that means it can be either phishy or legitimate, meaning the website held some legit and phishy features.

SHOW LESS ^

Has Missing Values?

No

Introductory Paper

Phishing detection based Associative Classification data mining

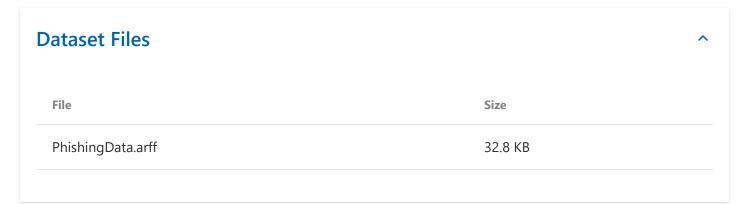
By Neda Abdelhamid, A. Ayesh, F. Thabtah. 2014 Published in Expert systems with applications

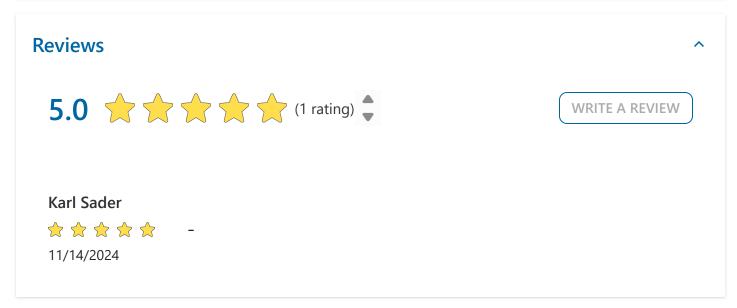
ariables Table					
Variable Name	Role	Туре	Description	Units	Missing Values
SFH	Feature	Integer			no
popUpWindow	Feature	Integer			no
SSLfinal_State	Feature	Integer			no
Request_URL	Feature	Integer			no
URL_of_Anchor	Feature	Integer			no
web_traffic	Feature	Integer			no
URL_Length	Feature	Integer			no
age_of_domain	Feature	Integer			no
having_IP_Address	Feature	Integer			no
Result	Target	Integer			no
		Rows po	er page 10) 0 to 10	of 10 〈 >

Additional Variable Information

URL Anchor Request URL SFH URL Length
Having '@'
Prefix/Suffix
IP
Sub Domain
Web traffic
Domain age
Class

collected features hold the categorical values , "Legitimateâ€□, â€□Suspiciousâ€□ and "Phishyâ€□, these values have been replaced with numerical values 1,0 and -1 respectively. details of each feature are mentioned in the research paper mentioned below SHOW LESS ^







CITE

- 99 1 citations
- **15100** views

Keywords

(phishing)

Creators

Neda Abdelhamid

DOI

10.24432/C5B301

License

This dataset is licensed under a **Creative Commons Attribution 4.0 International** (CC BY 4.0) license.

This allows for the sharing and adaptation of the datasets for any purpose, provided that the appropriate credit is given.

THE PROJECT

About Us

CML

National Science Foundation

NAVIGATION

Home

View Datasets

Donate a Dataset

LOGISTICS

Contact

Privacy Notice

Feature Request or Bug Report