UCI

Machine Learning Repository

Center for Machine Learning and Intelligent Systems

Search

○ Repository ○ Web

Google™

View ALL Data Sets

Website Phishing Data Set

Download: Data Folder, Data Set Description

Abstract:

Data Set Characteristics:	Multivariate	Number of Instances:	1353	Area:	Computer
Attribute Characteristics:	Integer	**Number of Attributes:**	10	**Date Donated**	2016-11-02
Associated Tasks:	Classification	**Missing Values?**	N/A	**Number of Web Hits:**	77723

Source:

Neda Abdelhamid
Auckland Institute of Studies
nedah '@' ais.ac.nz

Data Set Information:

The phishing problem is considered a vital issue in "COM" 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.

Attribute 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 , â€œLegitimateâ€, â€Suspiciousâ€ and â€œPhishyâ€, 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

Relevant Papers:

You can view all citations that used the paper that has applied this data, mentioned below at [Web Link]

Citation Request:

Abdelhamid et al.,(2014a) Phishing Detection based Associative Classification Data Mining. Expert Systems With Applications (ESWA), 41 (2014) 5948â€"5959.

Supported By: In Collaboration With: Rexa.info
Research • People • Connections

About || Citation Policy || Donation Policy || Contact || CML