



PhiUSIIL Phishing URL (Website)

Donated on 3/3/2024

PhiUSIIL Phishing URL Dataset is a substantial dataset comprising 134,850 legitimate and 100,945 phishing URLs. Most of the URLs we analyzed, while constructing the dataset, are the latest URLs. Features are extracted from the source code of the webpage and URL. Features such as CharContinuationRate, URLTitleMatchScore, URLCharProb, and TLDLegitimateProb are derived from existing features.



Dataset Characteristics

Tabular

Subject Area

Computer Science

Associated Tasks

Classification

Feature Type

Real, Categorical, Integer

Instances

235795

Features

54

Dataset Information



What do the instances in this dataset represent?

URLs and their corresponding webpages

Has Missing Values?

No

Introductory Paper



[PhiUSIIL: A diverse security profile empowered phishing URL detection framework based on similarity index and incremental learning](#)

By Arvind Prasad and Shalini Chandra. 2024

Published in Computers & Security

Variables Table



Variable Name	Role	Type	Description	Units	Missing Values
FILENAME	Other	Categorical			no
URL	Feature	Categorical			no
URLLength	Feature	Integer			no
Domain	Feature	Categorical			no
DomainLength	Feature	Integer			no

Rows per page

5

0 to 5 of 56



Additional Variable Information



Column "FILENAME" can be ignored.

Class Labels

Label 1 corresponds to a legitimate URL, label 0 to a phishing URL

Dataset Files



File

Size

PhiUSIIL_Phishing_URL_Dataset.csv

54.2 MB

Reviews



DOWNLOAD (14.7 MB)



IMPORT IN PYTHON

 1 citations 49128 views

Citations/Acknowledgements

If you use this dataset, please cite:

Prasad, A., & Chandra, S. (2023). PhiUSIIL: A diverse security profile empowered phishing URL detection framework based on similarity index and incremental learning. *Computers & Security*, 103545. doi: <https://doi.org/10.1016/j.cose.2023.103545>



Creators

Arvind Prasad

 arvindbitm@gmail.com

Babasaheb Bhimrao Ambedkar University

Shalini Chandra

 Babasaheb Bhimrao Ambedkar University

DOI

<https://doi.org/10.1016/j.cose.2023.103545>

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](#)