Deep learning based classification model to detect fake websites used in phishing attacks

Fabiano de M. Domingues[1], *Carlos T. P. Zanini*[1]
[1] Federal University of Rio de Janeiro

Abbreviated abstract: Phishing is a cybercrime that uses social engineering techniques to trick internet users into obtaining their confidential information. Usually, it is based on technological mechanisms that direct these users to fake pages in order to capture data such as passwords and credit card numbers. Deep learning techniques will be applied on a dataset containing extracted features from phishing and legitimate pages, resulting in a classification model that can be integrated with the mechanisms for detecting and blocking fake websites.

Related publications:

[1] Chiew, K. L. et al., Information Sciences 484 (C), 153-166 (2019).
[2] Tan, C. L., Mendeley Data 1 (2018)

UFRJ

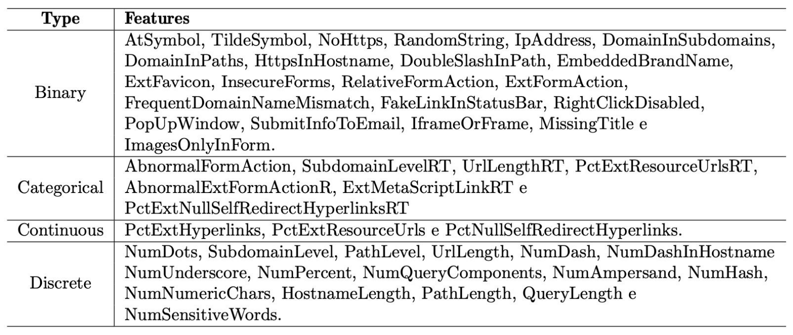
4th Conference on
Statistics and
Data Science
Salvador, Brazil (online)
December 1-3, 2022

Dataset, Previous Works and Challenge

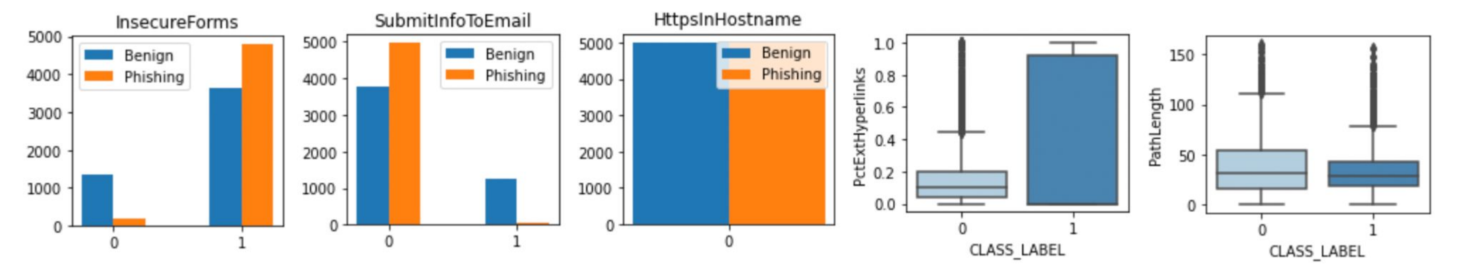
Dataset contains 48 features extracted from web page URLs and HTML codes, from 5,000 internet pages used in phishing attacks and 5,000 legitimate pages [1].

Previous work uses machine learning techniques: SVM, Naive Bayes, C4.5, Random Forest, JRip, and PART [2],

The challenge is to investigate the use of deep learning techniques to achieve better performance.



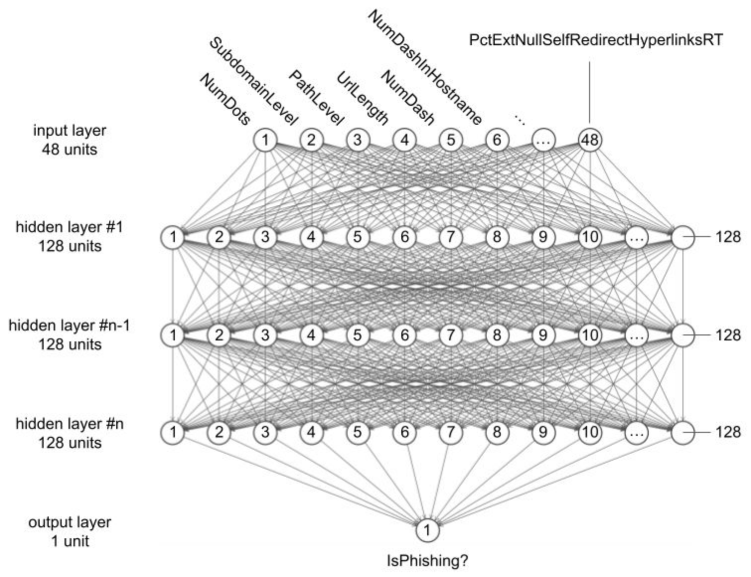
Type	Features
Binary	AtSymbol, TildeSymbol, NoHttps, RandomString, IpAddress, DomainInSubdomains, DomainInPaths, HttpsInHostname, DoubleSlashInPath, EmbeddedBrandName, ExtFavicon, InsecureForms, RelativeFormAction, ExtFormAction, FrequentDomainNameMismatch, FakeLinkInStatusBar, RightClickDisabled, PopUpWindow, SubmitInfoToEmail, IframeOrFrame, MissingTitle e ImagesOnlyInForm.
Categorical	AbnormalFormAction, SubdomainLevelRT, UrlLengthRT, PctExtResourceUrlsRT, AbnormalExtFormActionR, ExtMetaScriptLinkRT e PctExtNullSelfRedirectHyperlinksRT
Continuous	PctExtHyperlinks, PctExtResourceUrls e PctNullSelfRedirectHyperlinks.
Discrete	NumDots, SubdomainLevel, PathLevel, UrlLength, NumDash, NumDashInHostname NumUnderscore, NumPercent, NumQueryComponents, NumAmpersand, NumHash, NumNumericChars, HostnameLength, PathLength, QueryLength e NumSensitiveWords.



UFRJ

4th Conference on
Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

Methods



Multilayer Perceptron models (1 to 4 hidden layers)

A baseline model: Logistic regression

Removed *HttpsInHostname* feature after exploratory analysis (input layer from 48 to 47 neurons)

Extracted division of the dataset (10,000) into test (1,000), evaluation (900) and training (8,100).

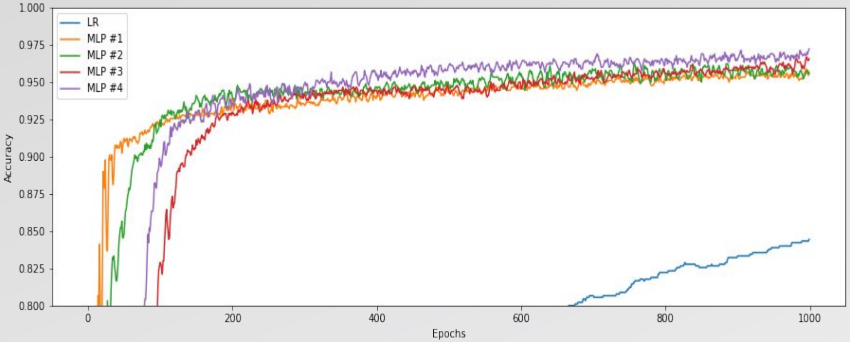
Best performance of each model is based on the highest accuracy in the evaluation set at one of the 100 training epochs.

The best performing MLP and baseline model are used in test phase.

UFRJ

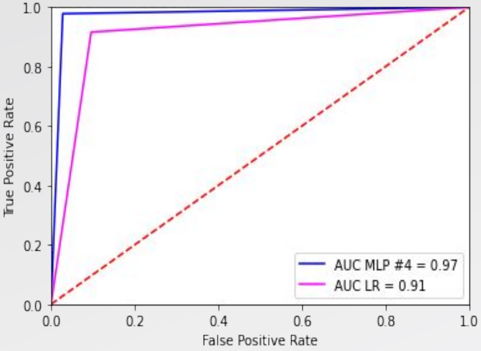
4th Conference on
Statistics and
Data Science
Salvador, Brazil (online)
December 1-3, 2022

Results and Conclusions



Model	Accuracy
LR	0.844
MLP #1	0.959
MLP #2	0.962
MLP #3	0.967
MLP #4	0.972

Model	Accuracy
LR	0.905
MLP #4	0.972



The MLP with 4 hidden layers showed better performance during training and evaluation, and relevant distance from the baseline model in test.

The increase in accuracy as new layers were added shows that it can be interesting to perform tests with 5 or 6 hidden layers.

Possibility of pairing experiments for comparison with previous work [2], which distinguishes 94.6% of sites using 20.8% of the original features with Random Forest.

UFRJ

4th Conference on
Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022