Recognizing phishing websites based on a bayesian combiner

Document Type : Research Paper

Authors

Omid Rahmani Seryasat ✉ [1] ; Sina Ahmadi [2] ; Pouya Yousefi [2] ; Farzad Tat Shahdost [3] ; Sareh Sanei [4]

[1] Department of Electrical Engineering, Shams Higher Education Institute, Iran.

[2] Department of Computer Engineering, West Tehran Branch, Islamic Azad University, Tehran, Iran.

[3] Department of Electrical Engineering, Islamic Azad University, Garmsar Branch, Semnan, Iran.

[4] Department of Electrical Engineering, Technical and Vocational University (TVU), Tehran, Iran.

Abstract

Phishing is a social engineering technique used to deceive users, which means trying to obtain confidential information such as username, password or bank account information. One of the most important challenges on the Internet today is the risk of phishing attack and Internet scams. These attacks cost the United States billions of dollars a year. Therefore, researchers have made great efforts to identify and combat such attacks. Accordingly, the present study aims to evaluate the methods of identifying phishing websites. This research is applied in terms of its objectives and descriptive-analytical in nature. In this article, the classification approach is used to identify phishing websites. From a machine learning point of view, if a suitable strategy is used, the ensemble of votes of different classifiers can be used to increase the accuracy of classification. In the method proposed in this paper, three inherently different ensemble classifiers, called bagging, AdaBoost, and rotation forest are employed. In this method, the stacked generalization strategy is used as an ensemble strategy. A relatively new dataset is employed to evaluate the performance of the proposed method. The database was added to the UCI Database in 2015 and uses 30 features that appear to be appropriate for distinguishing phishing and non-phishing websites. The present study uses 10-fold-cross-validation method as an evaluation strategy. The numerical results indicate that the proposed method can be used as a promising method for detecting phishing websites. It is worth mentioning that in this method, an F-score of 96.3 is resulted, which is a good result in detecting phishing.

Keywords

Phishing ; Classification ; Ensembling ; Stacked generalization

Files

XML

PDF 455.18 K

Share

How to cite

Statistics

Article View: *24*

PDF Download: *4*