

Assignment

(100 Points)

CYBR 7240

Cyber Analytics and Intelligence

Module 02: Linear & Logistic regression

In this homework, we explore “Phishing Dataset” using logistic regression.

- 1) (50 points) Apply logistic regression technique using WEKA tool in 4 different settings including:
 - 5 fold-cross validation.
 - 10 fold-cross validation.
 - 15 fold-cross validation.
 - 80% training.
- 2) (50 points) Remove 10 features randomly from the dataset and repeat the same procedure (part 1) again.
- 3) (25 points) Bonus: do some research on Feature Selection in Machine Learning (ML) and briefly explain why it is an important task in ML. Additionally, try to apply different feature selection techniques on the dataset and see if you can improve the accuracy of part 1.
(hint: <https://www.youtube.com/watch?v=x5wa1w-BpRE&t=277s>)

Write your report and compare the results.

Deliverable:

- Your report including the screenshots of your implementation and the result.