

Classification Project Proposal

Jamila Rabeh Alharbi

Ghalia Maher

Ghadeer Ali Alghamdi

-
- Question/Problem statement/Goals:

“Is this a trusted link?”

In today’s digital age, and with the digital transformation that we are witnessing nowadays not locally in our country only but in the whole world, cybercrime affects all of us directly or indirectly because as individuals and companies, we all have information that is worth something to cybercriminals. The most cybercrime common way to attack an individual is phishing websites especially lately due to COVID-19, the use of the internet and E-commerce, E-governance sites has increased.

Dr.Khalid Bin Abdullah Alsabti the governor of the Cybersecurity Authority has announced on 7th April of 2021 in the Opening of the global cybersecurity conference that phishing sites increased by 300%! This is a huge increase.

Therefore, in this project we will determine whether this link is phishing or not by applying multi classifier methods and evaluate them using classifications metrics.

- Data Description:
 - Dataset published date: 2019
 - Dataset source:
<https://www.sciencedirect.com/science/article/pii/S2352340920313202>
 - Data Size: It has 112 columns (Features) and 88,648 rows (Records).
 - Data type: (Integer,Float)
- Tools:
 - Programs: Python using Jupyter notebook.
 - Libraries: Pandas, Numpy, sklearn.
 - Functions: .median .count .max .mean
 - Plots: Bar Charts, Pie Charts
 - Methods: Logistic regression, Random Forest classifier and naive_bayes