Phishing Websites Detector for CS 235

Yash Aggarwal
UC Riverside
yagga004@ucr.edu

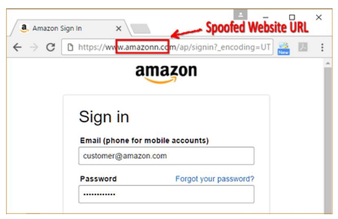
Hritvik Gupta
UC Riverside
hgupt010@ucr.edu

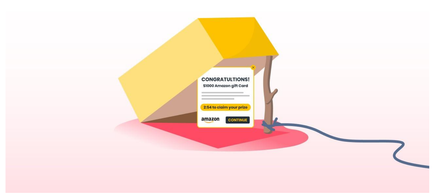
Siddhant Poojary
UC Riverside
spooj003@ucr.edu

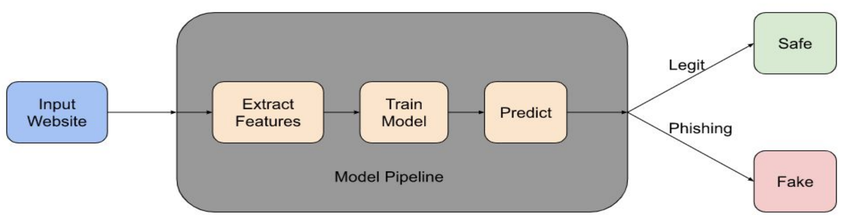
Nityash Gautam
UC Riverside
ngaut006@ucr.edu

Shubham Sharma
UC Riverside
sshar180@ucr.edu

Introduction



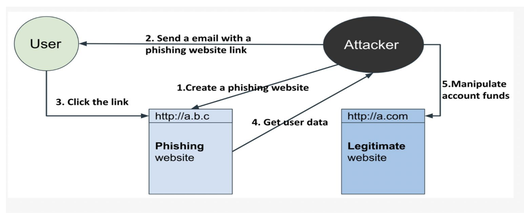




What is Phishing?
- Spoofed website to trick user to reveal sensitive information to attacker
- Most Common Cyber attack as of 2022

Scope of this work
- Given an input website, Extract useful features and use them to predict if the website is legit or phishing

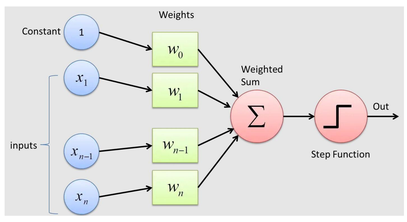
Problem Definition



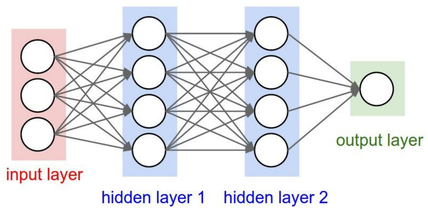
What's the problem?
- Phishing website imitate the legitimate and trick user
- Hack into the system
- Need to distinguish those sites with correctness
- Machine learning classifiers can help to achieve those

Proposed Method

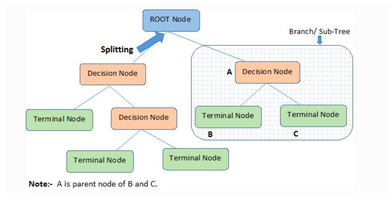
We will be attempting to classify websites into legit and benign using multiple algorithms like



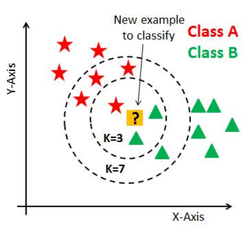
Perceptron



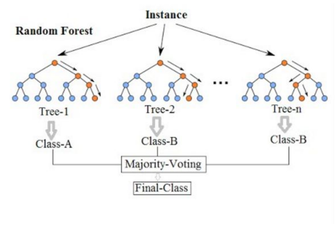
Multi layer Perceptron



Decision Tree



K-Nearest Neighbour



Random Forest

Results

Perceptron

METRICS	Baseline Implementation	Self Implementation
Precision	0.99	0.95
Recall	0.82	0.90
F1-Score	0.90	0.92
Miss Rate	0.013	0.05

Random Forest

METRICS	Baseline Implementation	Self Implementation
Precision	0.94	0.96
Recall	0.95	0.97
F1-Score	0.94	0.97
Miss Rate	0.04	0.02

MultiLayer Perceptron

METRICS	Baseline Implementation	Self Implementation
Precision	0.9	0.89
Recall	0.9	0.89
F1-Score	0.89	0.88
Miss Rate	0.09	0.13

Decision Trees

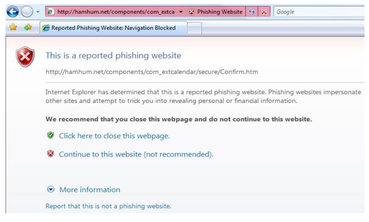
METRICS	Baseline Implementation	Self Implementation
Precision	0.94	0.95
Recall	0.94	0.95
F1-Score	0.94	0.95
Miss Rate	0.05	0.06

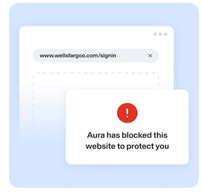
k-Nearest Neighbor

METRICS	Baseline Implementation	Self Implementation
Precision	0.95	0.95
Recall	0.94	0.95
F1-Score	0.95	0.95
Miss Rate	0.05	0.05

Related Work

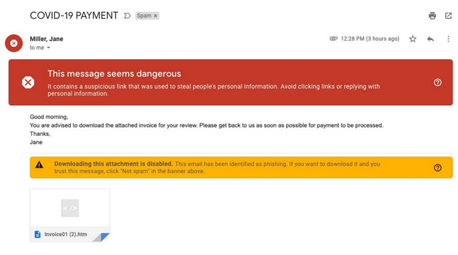
Similar works have been done by web browsers to classify websites as phishing based on its characteristics

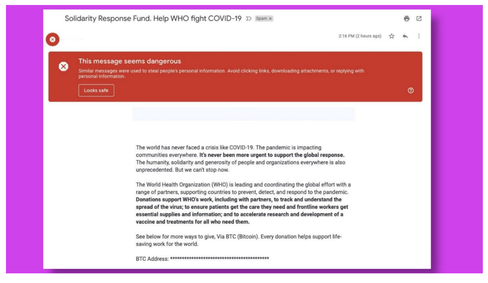




Get Alerted to Dangerous Sites
Aura automatically stops you from entering malicious sites, like malware and phishing sites, that may steal your personal and financial info.

Another related aspect if the emails that we receive might be phishing and gmail detects and marks them as phishing





Conclusions

Algorithms Implemented - Baseline and Self
i. Perceptron
ii. MultiLayer Perceptron
iii. Random Forest
iv. k-Nearest Neighbor
v. Decision Trees

Performance Evaluation Metric – ***MISS RATE***
Baseline Implementation - Best Performance - ***PERCEPTRON***
Self Implementation - Best Performance - ***RANDOM FOREST***