WEAK PASSWORD, MALICIOUS URL & PHISHING URL DETECTION

TABLE OF CONTENT

INFORMATION SECURITY AUDIT AND ANALYSIS -CSE 3061

TEAM MEMBERS:

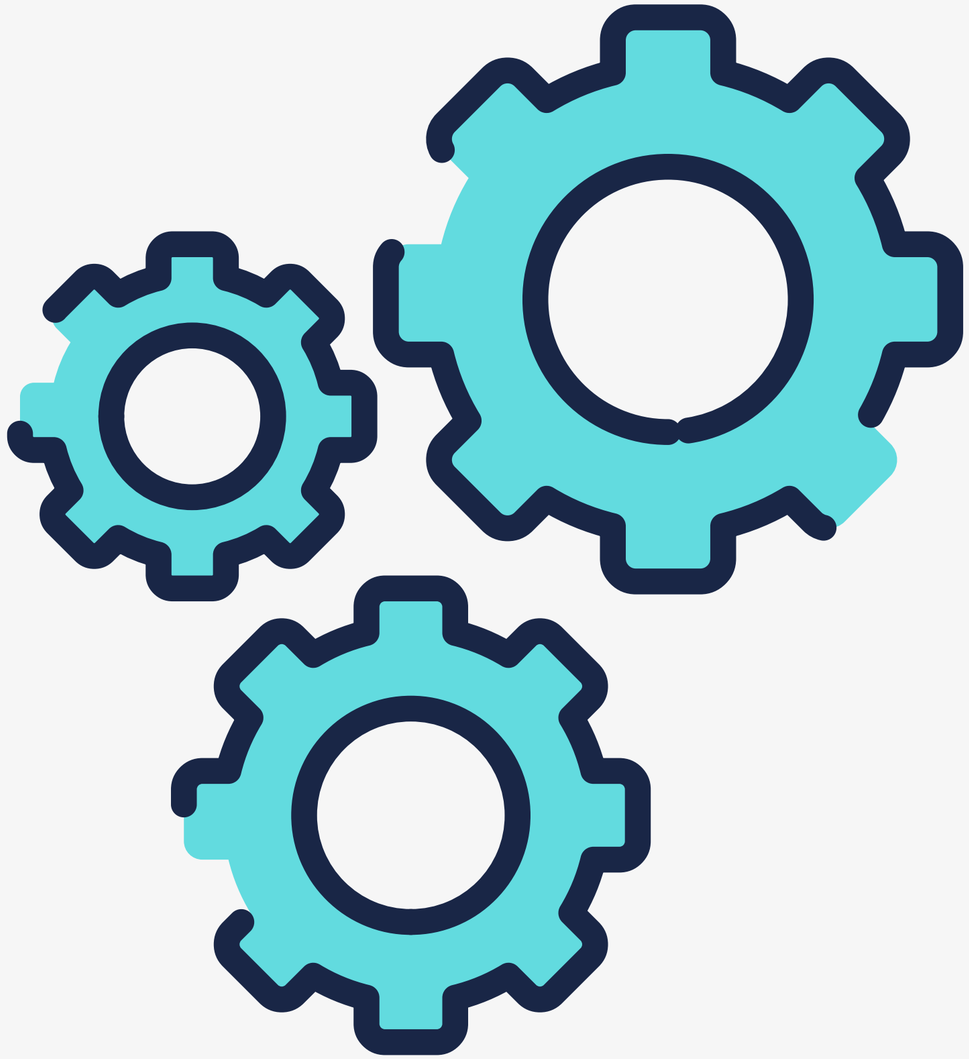
20BAI1026 - SUBRITTA CHATTERJEE

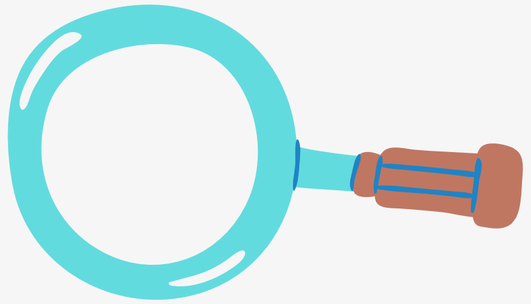
20BAI1085 - JAYANAND JAYAN

20BAI1213 - ROHITA CHAKRABORTY

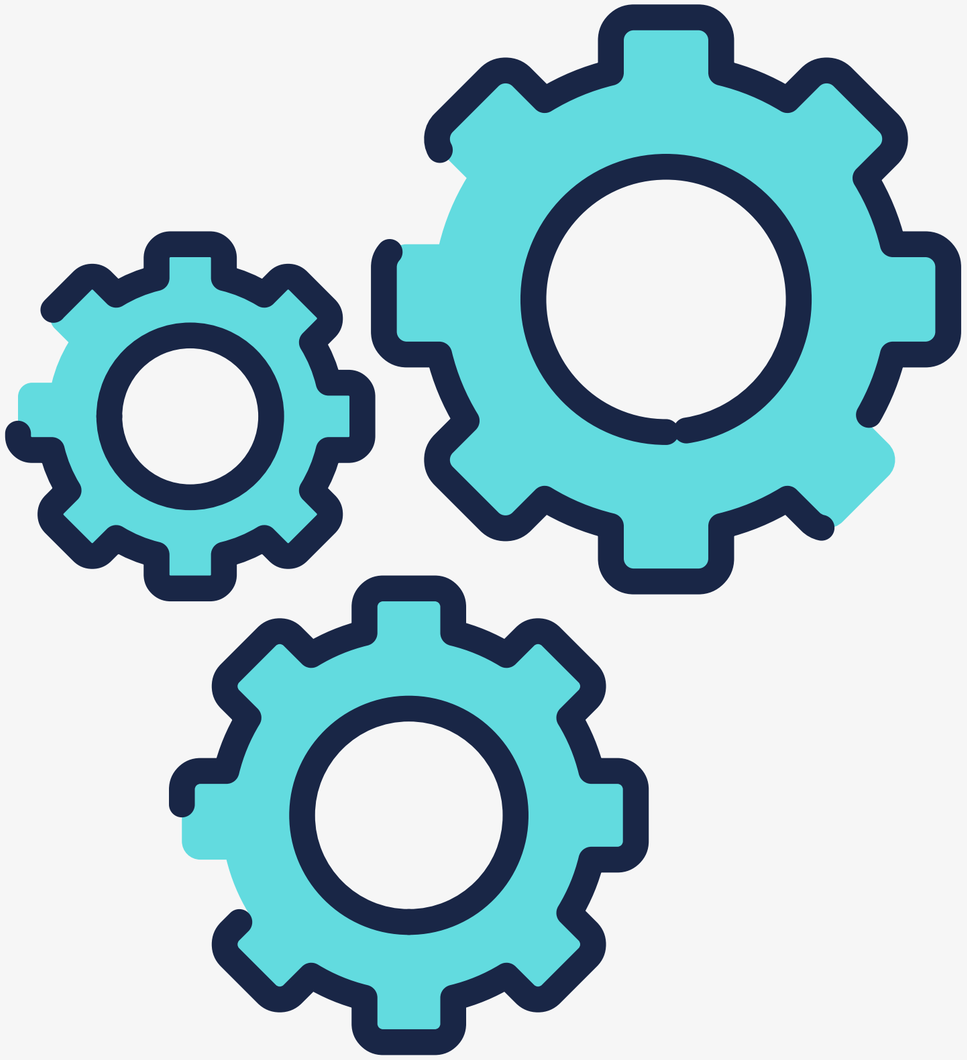
GUIDE NAME - RUKMANI P

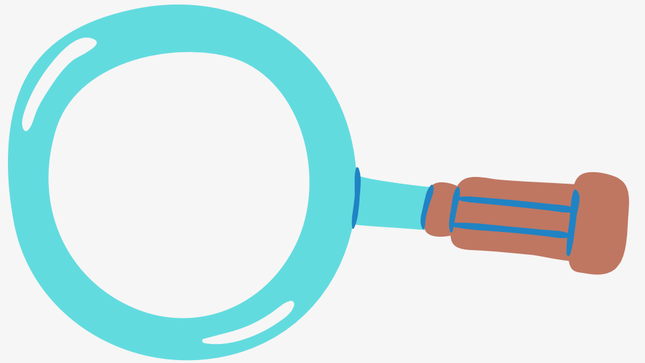
ABSTRACT & PROBLEM STATEMENT



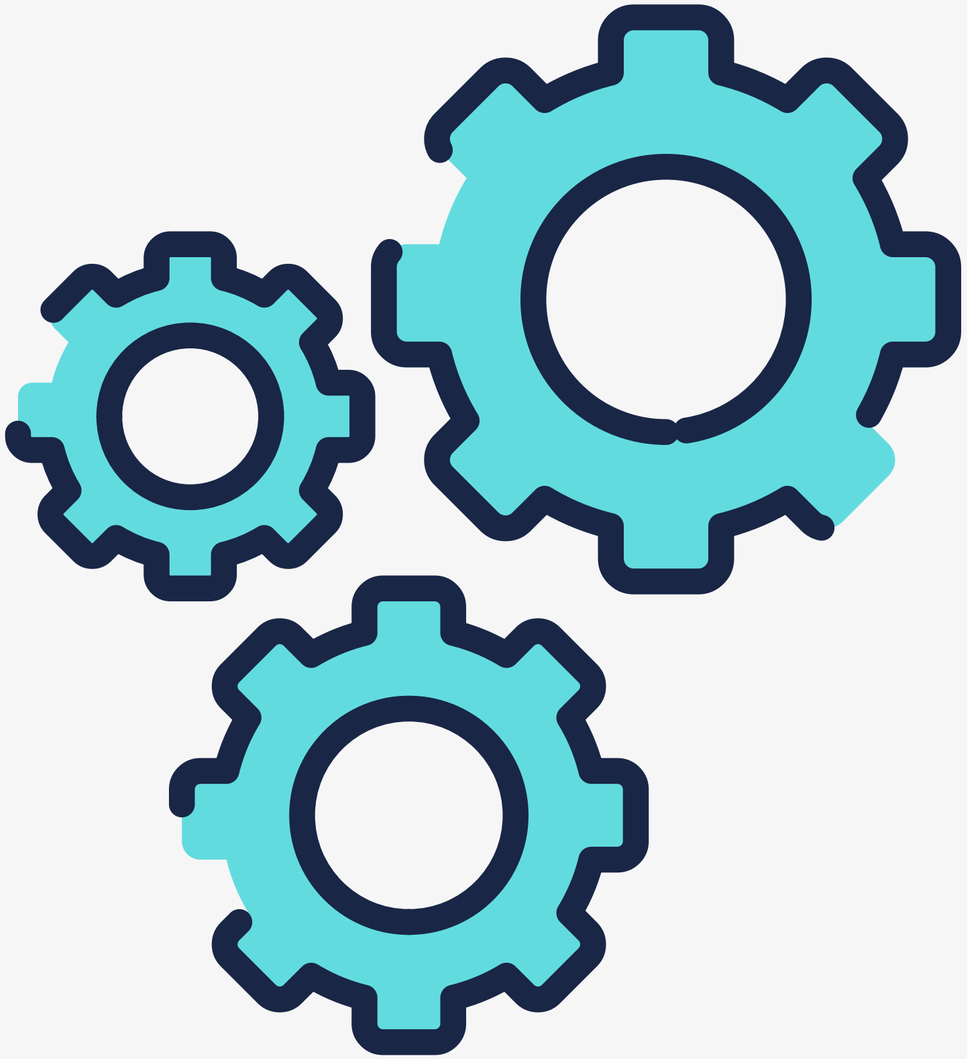


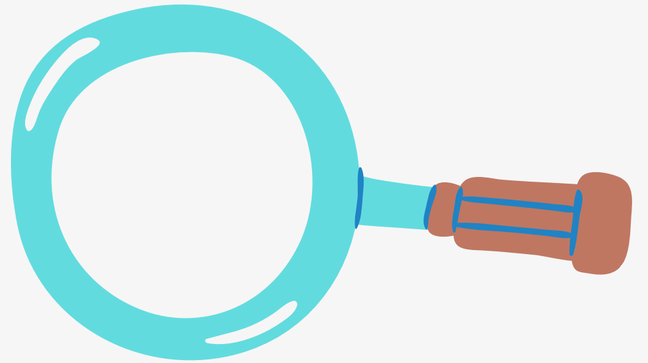
Weak passwords always play a major role in any hack. For the ease of user, sometime applications do not enforce password complexity and as a result of that users use simple passwords. Weak passwords can be guessable or attacker can brute force if the length of the password is very small, so try to use random strings with special characters.





Attackers often try to change one or more components of the URL's structure to deceive users for spreading their malicious URL. **Malicious URLs** are known as links that adversely affect users. These URLs will redirect users to resources or pages on which attackers can execute codes on users' computers, redirect users to unwanted sites, malicious website, or other phishing site, or malware download. Malicious URLs can also be hidden in download links that are deemed safe and can spread quickly through file and message sharing in shared networks





A **phishing website** is a common social engineering method that mimics trustful uniform resource locators (URLs) and webpages. Through phishing attacks, the phisher targets naïve online users by tricking them into revealing confidential information, with the purpose of using it fraudulently. Main aim of the attacker is to steal banks account credentials. In United States businesses, there is a loss of US$2billion per year because their clients become victim to phishing.

DATASET

We will collect the dataset for **weak passwords** from a 000webhost leak, which is available online. The commercial password strengths algorithms used are of Twitter, Microsoft and battle. A password in the dataset is classified into a category only if it was categorized into the same category by all three algorithms.

We will collect the **malicious URLs** dataset from the ISCX-URL-2016 URL dataset. To boost the amount of data available, we will also combine the dataset with malware URLs from the Malware URL dataset blacklist. We will also collect benign URLs while making the model.

The set of **phishing URLs** are collected from opensource service called PhishTank. This service provides a set of phishing URLs in multiple formats like csv, json etc. that gets updated hourly.

The legitimate URLs are obtained from the open datasets of the University of New Brunswick, This dataset has a collection of benign, spam, phishing, malware & defacement URLs. Out of all these types, we will consider the benign url dataset.

OUR WORKFLOW

- Collect dataset containing weak passwords, malicious URLs, phishing and legitimate websites.
- Write a code to extract the required features
- Analyze and preprocess the dataset.
- Divide the dataset into training and testing sets.
- Run selected machine learning and deep neural network algorithms like SVM, Random Forest, MLP on the dataset.
- Display the evaluation results considering accuracy metrics.

- Compare the obtained results for trained models and specify which is better.
- Create a webpage with a user friendly GUI
- Deploy the machine learning model as an API
- Take the user input for a password or a URL based on the user's choice of service.
- Make an API call in the backend to the deployed machine learning model.
- Return the output and display it to the user.


Thank you!