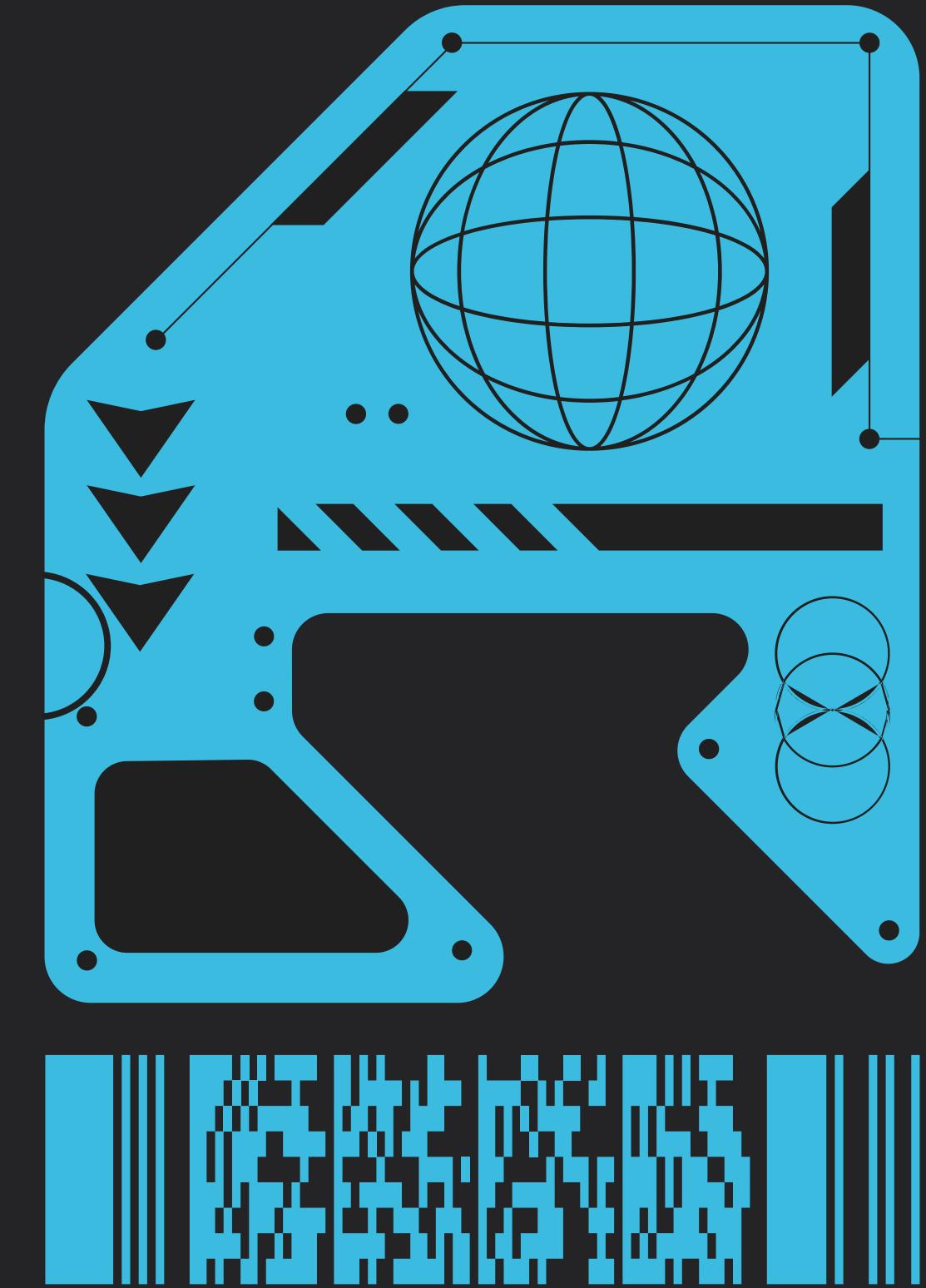


NOT SO PHISHY

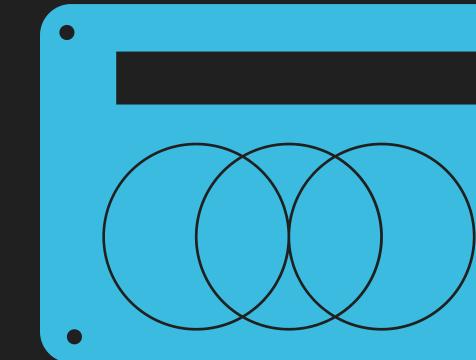


BRIEF

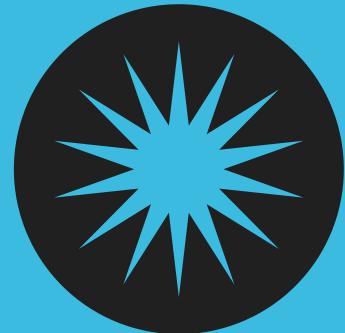
An ML based Real-Time URL classifier that identifies benign and malicious URLs based on attributes of the URL after passing them through the model.



TEAM TECH FUSION 



MEET THE TEAM



AKSHAY NEGI
E22CSEU0264

Contributed in Model
Training and Feature
Extraction

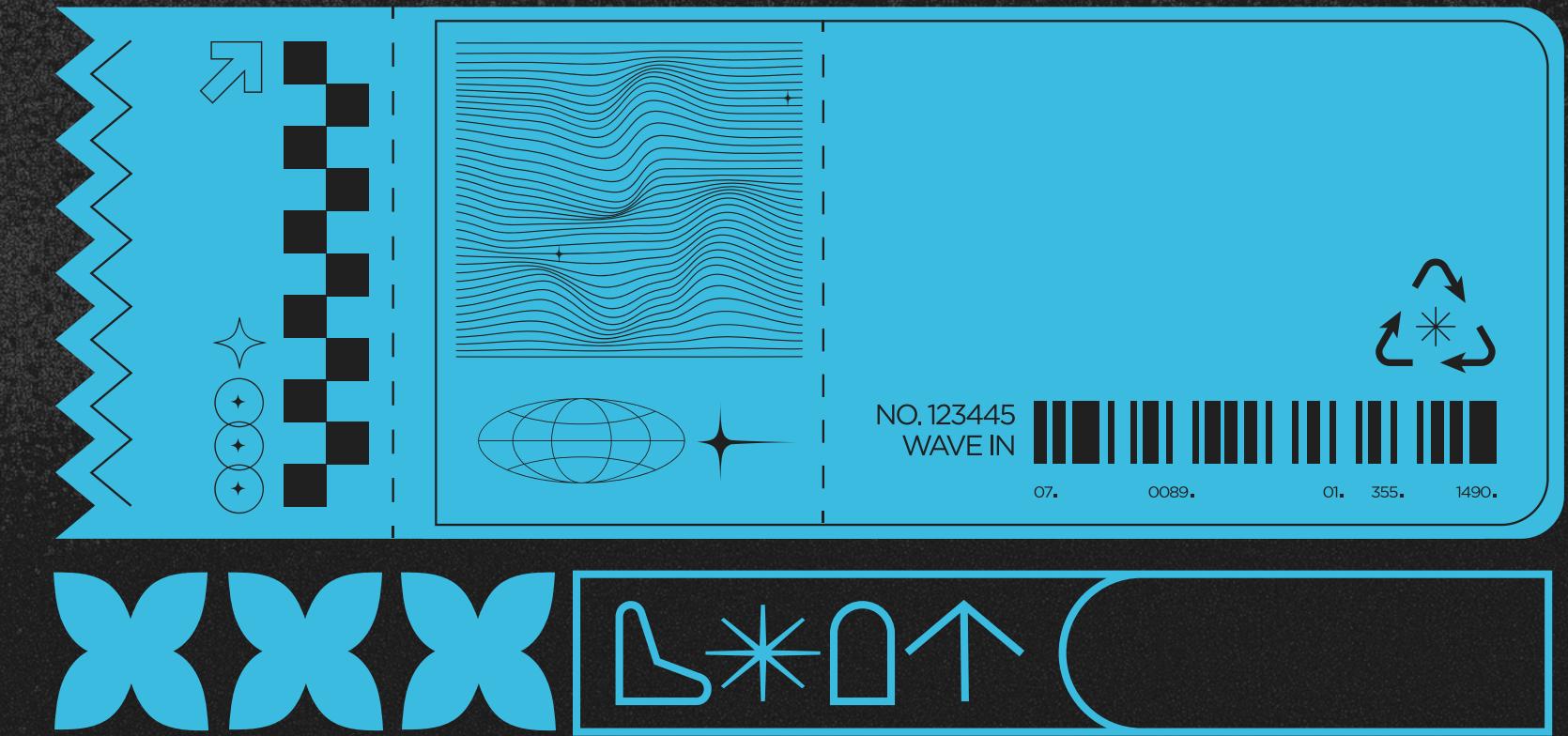
ADITI CHATURVEDI
E22CSEU0252

Contributed in Model
Training and Backend
Integration

SATYAM RAJ
E22CSEU0257

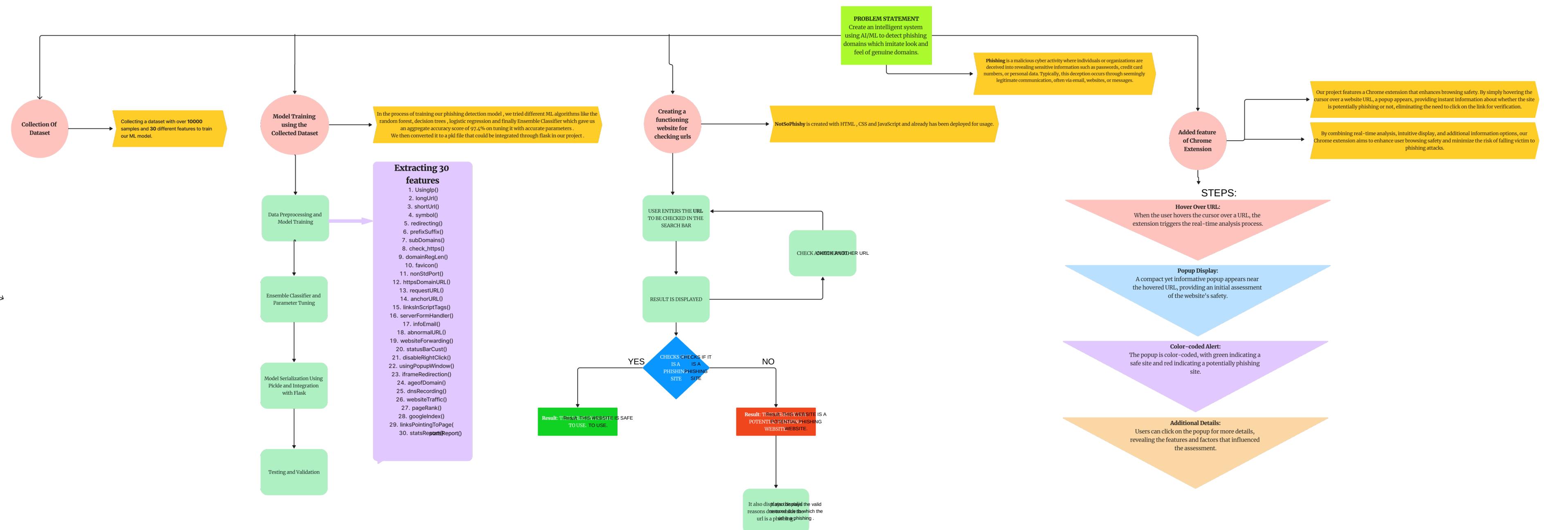
Contributed in Model
Training and Frontend
Development

PROJECT DESCRIPTION



- Objective: Develop an AI-ML model to detect phishing websites using a dataset of 32+ parameters.
- Dataset: Consists of 11,000+ labeled samples of phishing and legitimate websites.
- Parameters: Include URL length, age of domain, SSL certificate, etc.
- Approach: Use machine learning algorithms (e.g., logistic regression, decision trees) for training the model.
- Feature Selection: Apply techniques like information gain and chi-square to select relevant parameters.
- Preprocessing: Handle missing values, normalize features, ensure data quality.
- Progress: Collected and cleaned dataset, split into training and testing sets.
- Results: Initial experiments will show model accuracy of 96% on the validation set.

FLOWCHART



IDEAS/ APPROACH

- Phishing is the most common cyber threat today, exploiting user vulnerability rather than system weaknesses.
- Social engineering, relying on human error, is more potent than hacking methods.
- Our solution involves providing users with tools to make better decisions, using systems to identify complex patterns.
- An ML model trained on 11,000+ samples predicts URL reliability to detect phishing sites before user interaction.
- Our website offers a list of identified phishing URLs, global attack reports, and a search feature for users to check suspicious URLs.

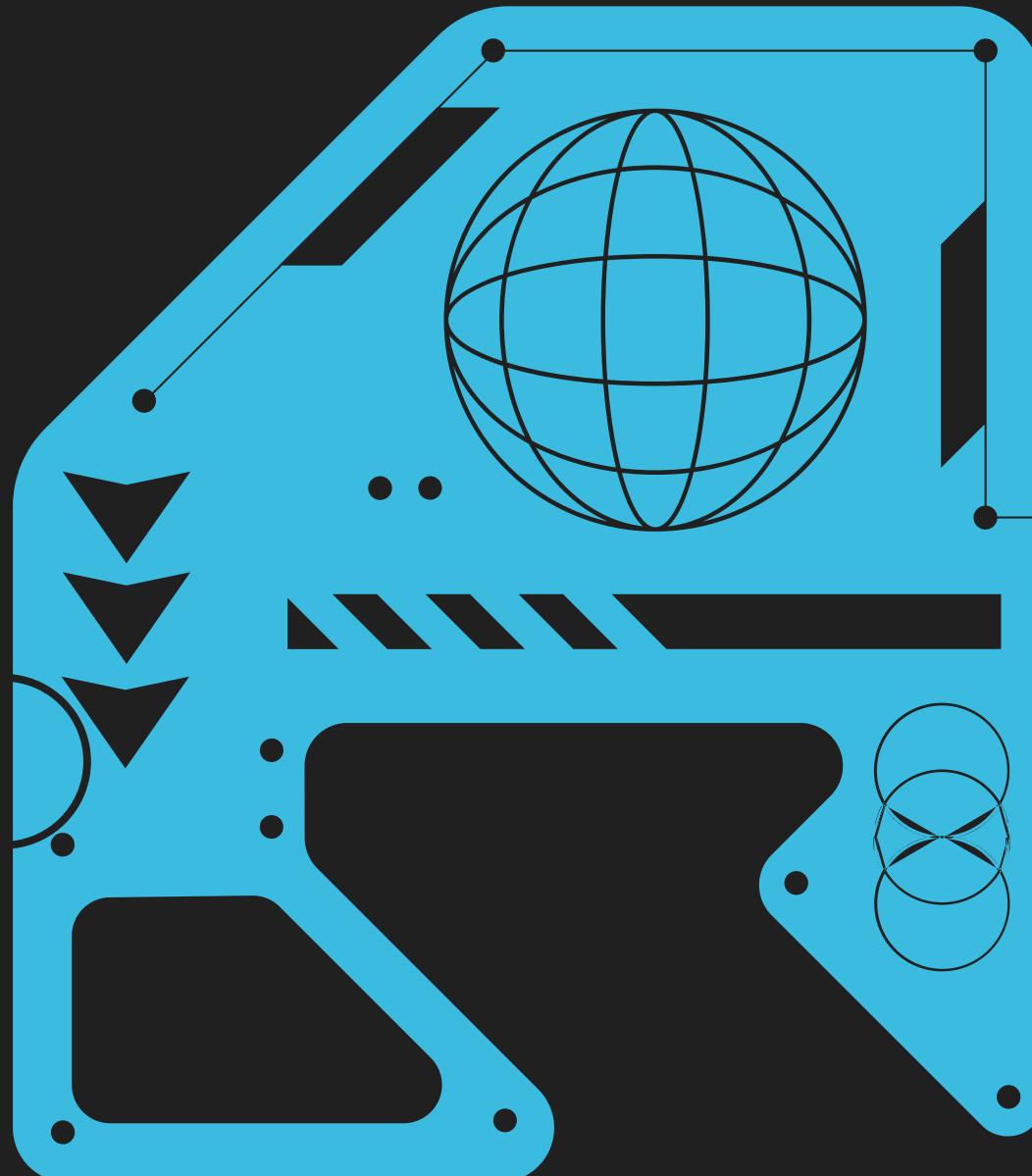


DEPENDENCIES AND NOVELTY

- We are focusing on including the entire functionality of this project as a google extension to increase the ease of use of the project.
- Our aim with the website is for it to check the website in realtime by checking all the parameters used by our model.
- Also, we are planning to flag and report the links identified as phishing URLs.

Dependencies:-

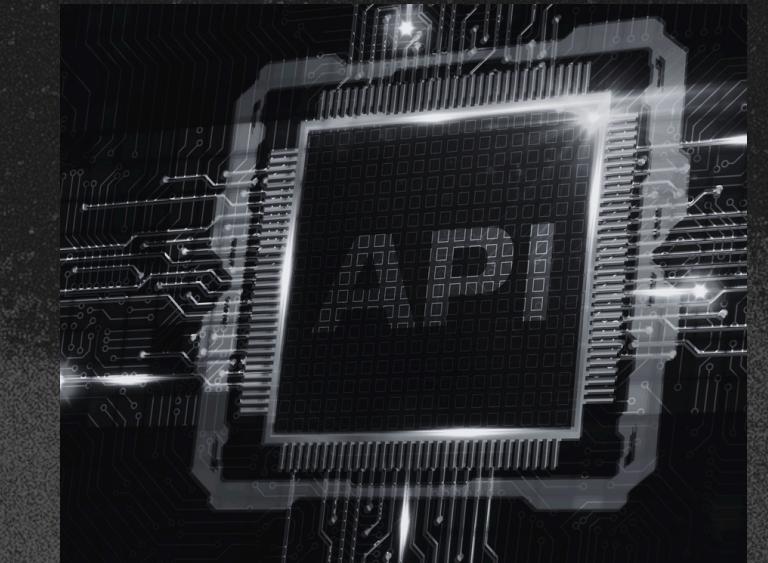
- A large set of data to train the AI/ML model.
- A way to collect and process data from the web.
- A way to connect to other systems and services.



SERVICES WE WILL PROVIDE.



Real-time Phishing Website Detection



Access to our ML model through API



Google Chrome Extension



D2C FOCUSED APPLICATION

WORK PROGRESS

DATA CLEANING ↴

FEATURE EXTRACTION ↴

WEBSITE DEPLOY ↴

STATUS

COMPLETED

STATUS

COMPLETED

STATUS

COMPLETED

We have completely cleaned the data, filled the missing values and preprocessed it completely to use it to train our model.

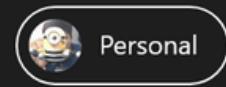
We have completely developed functions for extracting all the features that will be extracted from the website that has to be tested by our website.

We are now in process to train model and connect it to our website through flask so users can access the prototype to check whether the website is phishe or not.

**NOT SO PHISHY**[About Us](#)[Contact Us](#)[Get Extension](#)

Search for information of **phishing website connected to the public Internet.**





Personal



Phishing Analysis Result



127.0.0.1:5000/analyze

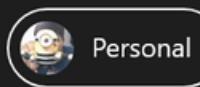


NOT SO PHISHY

[Get Extension](#)

This website is safe to use.

[Check Another URL](#)



Personal



Phishing Analysis Res



NOT SO PHISHY

Get Extension

Warning: The website may be phishing due to the following reason(s):

This URL has prefix or suffix added to the domain name.

The domain name in the URL doesn't match the domain name in the HTTPS certificate.

This URL loads content from different domains.

There are links in the HTML anchor tags pointing to external domains.

The URL has email addresses in it.

This URL deviates from typical URL structures.

This website customizes or disables the browser's status bar.

This website disables the right-click context menu.

This website uses pop-up windows.

This website has low traffic.

Journal of Health Politics, Policy and Law, Vol. 35, No. 4, December 2010
DOI 10.1215/03616878-35-4 © 2010 by The University of Chicago

Digitized by srujanika@gmail.com

The website has a huge number of external links pointing to the website.

Check Another URL

THANK YOU