HACKAGEEK - RGIPT

TEAM LINK - SHIELD

OBJECTIVE

The project aims to create a web extension that boosts user security by detecting potentially harmful links on webpages through machine learning models trained on a dataset. It visually notifies users by highlighting suspicious links without needing their immediate input, utilizing a real-time machine learning model to classify links.

COMPONENTS

Frontend: Code in HTML, CSS, and JavaScript that engages with the user and displays visual indicators.
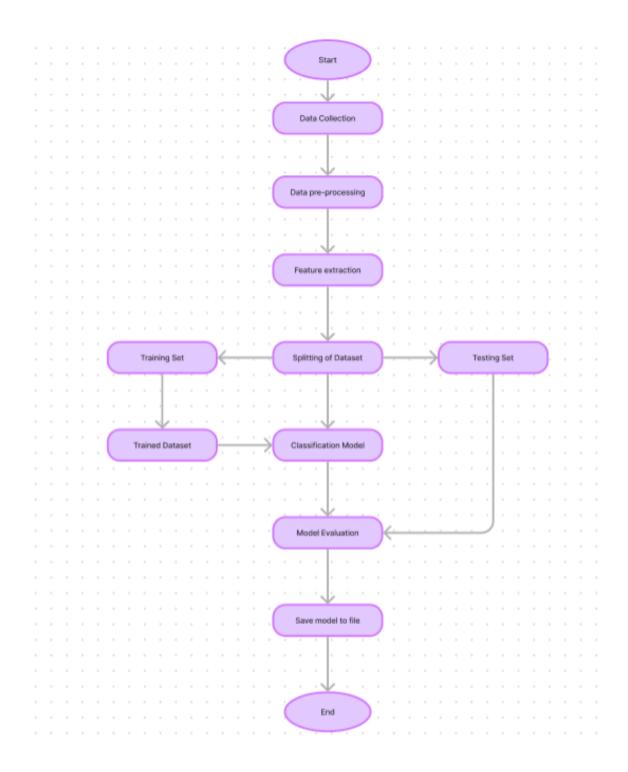
Backend: Integrating a web extension with a machine learning model via an API to analyse URLs. Using JavaScript along with Node.js and the Express framework to develop the API.

Machine Learning Model: A trained model that classifies URLs as phishing and legitimate. This model is trained using TensorFlow, a machine learning software library that is free and open-source. The dataset used can be found at https://github.com/goodycy3/Detection-of-Phishing-Website-Using-Machine-Learning/tree/master/ML%20work/DataSets. It consists of 5000 phishing links and 5000 legitimate links, in total 1000 links. These links are categorized into training and testing datasets after data preprocessing.

For training the model, following steps are to be followed -

1- Data Collection

2- Data Pre-processing

3- Data Analysis : Feature extraction

4- Splitting of dataset into training and testing dataset

5- Model Training on trained dataset using classification model

6- Model evaluation using testing dataset

7- Saving the model

8- Model deployment

Flow-Chart

Start

Data Collection

Data pre-processing

Feature extraction

Training Set ← Splitting of Dataset → Testing Set

Trained Dataset → Classification Model

Model Evaluation

Save model to file

End

WORKFLOW

URL Collection: The extension collects URLs from the webpages the user visits

URL Analysis: The data is sent to a machine learning model via an API to analyse each URL based on its features, such as the domain of the URL, presence of '@' symbol, 'http/https' in the domain name, redirection '//', and more.

Classification: The machine learning model categorizes URL as either phishing or legitimate.

Highlighting: If a URL is identified as phishing, the extension visually emphasizes it on the webpage through CSS styles.

Basic working