System Analysis

This chapter covers the feasibility study, requirements specification, system constraints and the use case diagrams for the proposed system.

4.1. Feasibility Study

This section covers the technical, economic, operational and social feasibility aspects of the proposed solution. This is necessary in order to assess the viability, practicality and effectiveness of developing the proposed solution

4.1.1. Technical Feasibility Study

Considering the hardware and software facilities required to develop and deploy the system, the platform is technically feasible. The hardware and software resources required to develop the system are immediately available and easily accessible. The infrastructure needed to host and run the system are available on cloud platforms.

The technologies required to develop the Machine Learning model are stable, well documented, and have a large community support to provide consultations in case of enquiries.

Use of cloud infrastructures facilitates the quick scaling, both up and down, of computing resources whenever needed to cater for the fluctuating number of user requests.

4.1.2. Economic Feasibility Study

The use of cloud computing services to host and run the system reduces the maintenance costs substantially. The only limiting factor is the cost incurred to the cloud service provider in order to host the system on a cloud premises.

The softwares and technologies required to develop the system are freely available and accessible on the internet.

Due to the high automation in the system, it will require very little labor to run and maintain. This ensures that there are minimal costs in securing human resources.

4.1.3. Operational Feasibility Study

With the use of cloud technology, the requirement to maintain the infrastructure upon which the system is deployed is outsourced to the cloud provider. This reduces the costs of maintenance of the system substantially. The system requires minimal human intervention in order to operate smoothly. Deploying the system on a cloud platform guarantees that the product will be available at any time over the internet. Cloud platforms also facilitate smooth transitions when upgrading to newer versions of the system.

4.1.4. Social Feasibility Study

The sole purpose of this system is to automate phishing websites detection for end users with minimal user input. This system will alert users in the event of potential threats thereby providing real time protection on every URL visit. Users will easily be able to integrate the system onto the supported browsers therefore reducing the complexity of using the system. Users will also easily adapt to the system.

This solution ensures a viable, beneficial and acceptable solution to the users therefore guaranteeing that the system is socially feasible.

4.2. Requirements Specification

4.2.1. Functional Requirements

1. The system should be able to classify a URL as legitimate or fraudulent based on the training dataset.

2. The system should be able to train the model based on the given data set.

3. The system should alert users when they visit a URL that identifies as a potential phishing website on their browser.

4. The system should allow users to input URLS, scan for their authenticity and finally give the appropriate feedack.

5. The system should keep track of URL lookups performed by users.

6. The system should have a commenting sections on the online platforms encourage users to interact.

7. The system should facilitate user account management in order to keep track of user contributions to the platform.

8. The system should be indexed by google search.

4.2.2. Non-Functional/Quality Requirements.

1. The system should be accessible at all times over the internet.

2. The system should be accessible across different devices as a website.

3. The application should be easily installable as a browser extension.

4. The system should be able to scale resources with increasing traffic.

5. The system should be secure to prevent stealing user information.

5. The application should not take not more than 10 seconds to alert users when a potential threat is detected.

6. The user interface should be able to scale across different screen sizes.

7. It should be easy to perfom URL scans on the website

8. The system should keep an audit trail of unsuccessful user requests.

4.2.3. Hardware Requirements

The following are the hardware requirements for developing the proposed system:

\* Processor - 500MHz and above

\* RAM - 8GB

\* Hard Disk - 500GB (at least 1 GB reserved for this project)

\* Input Devices (Standard Mouse, Keyboard)

4.2.4 Software Requiremnts

The following are the hardware requirements for developing the proposed system:

\* OS - Linux / Windows

\* Platform : Jupyter Notebook

\* Language : Python

\* IDE/Tool : Anaconda

4.3. System Constraints

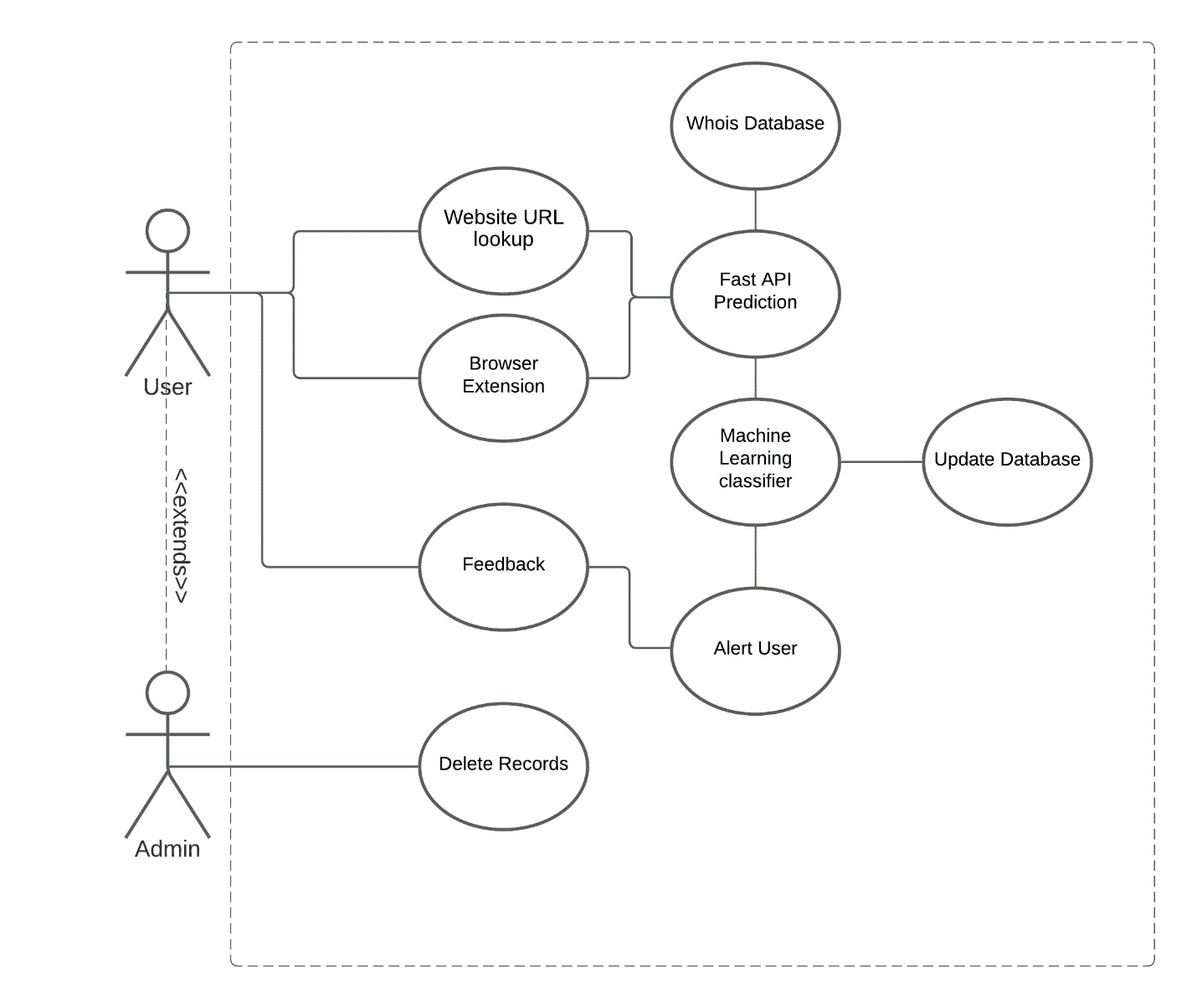
The system is limited to the following constraints:

\* In order to access the system, the end user is required to have a reliable internet connection to facilitate communication across the various interoperating services.

\* The browser extension made will be based on the chromium browser standardards hence it might not be compartible with every browser.

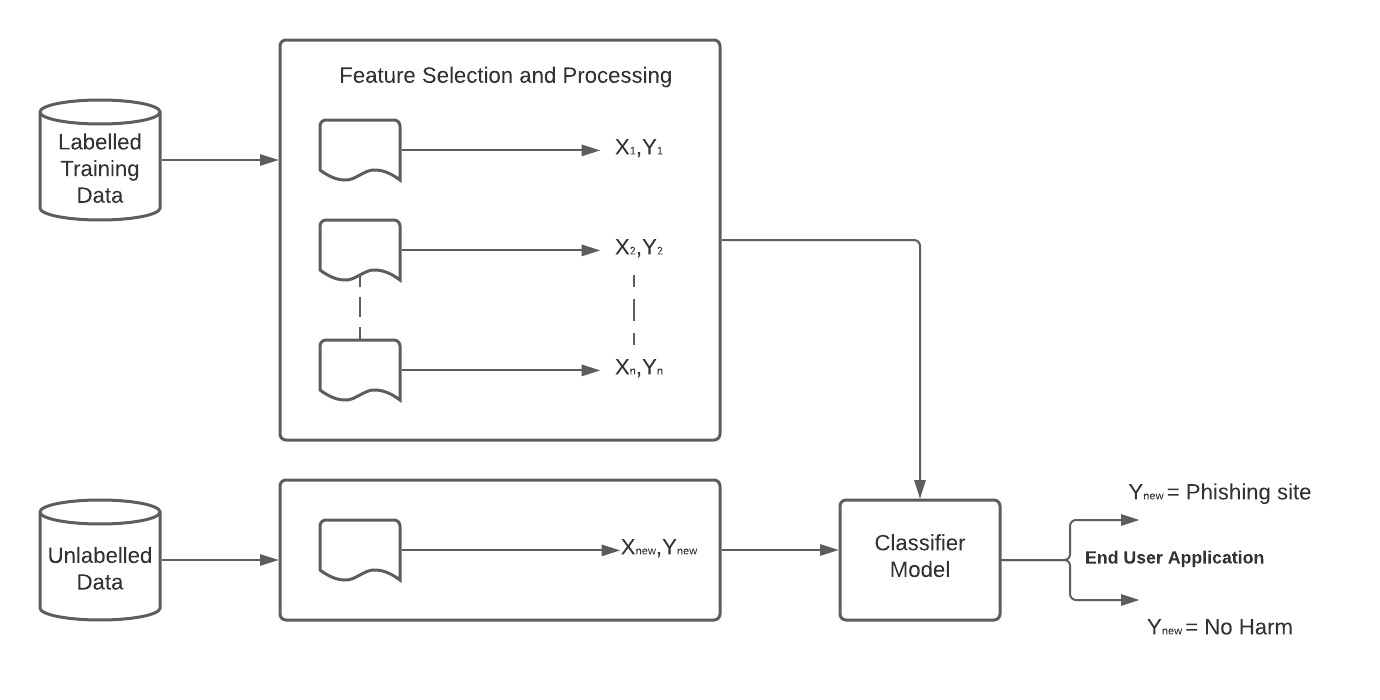
\* Android users will be required to use a specific subset of browsers in order to integrate the browser extension to their smartphones/tablets

4.4. Use Case Diagram



5.1. Conceptual Design

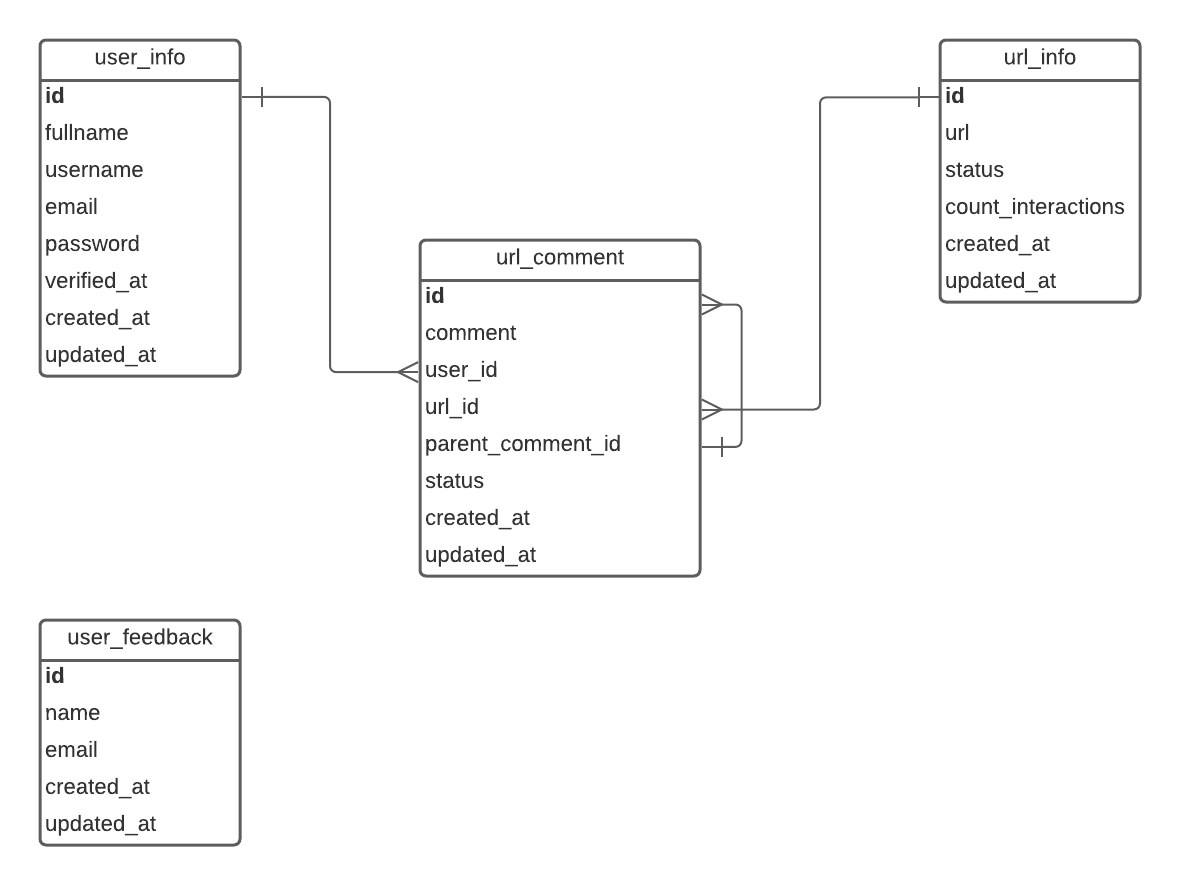
The proposed algorithm follows a supervised learning model architecture. Labelled data (URLs) will first be loaded into the system. The favorable features will be extracted and processed based on the nature of the data used. The extracted features then formed the basis for the classifier model. The unlabeled data (new URLs) will then be loaded to the classifier model for classification. Finally, the new data will be classified as either malicious or benign.



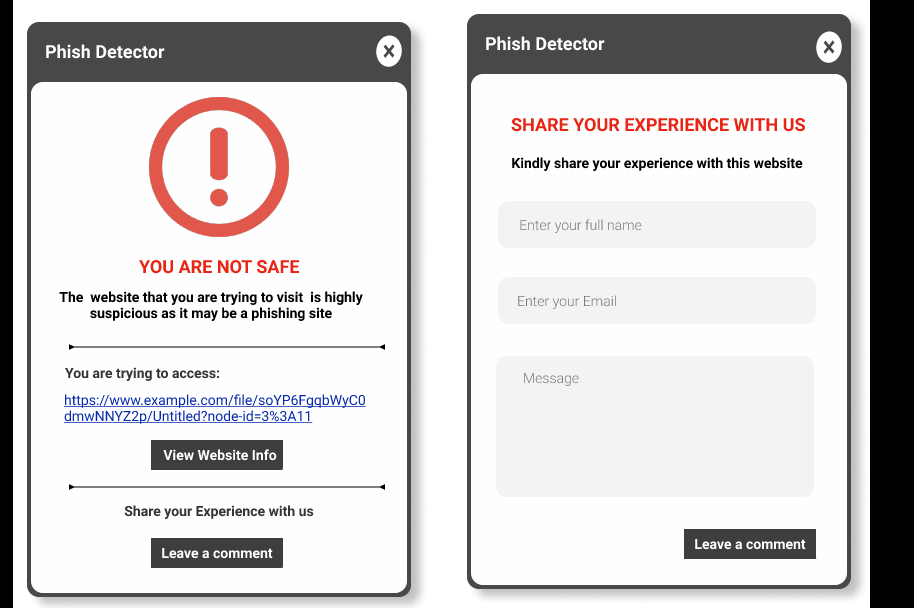
5.2. Architectural Design and System Components

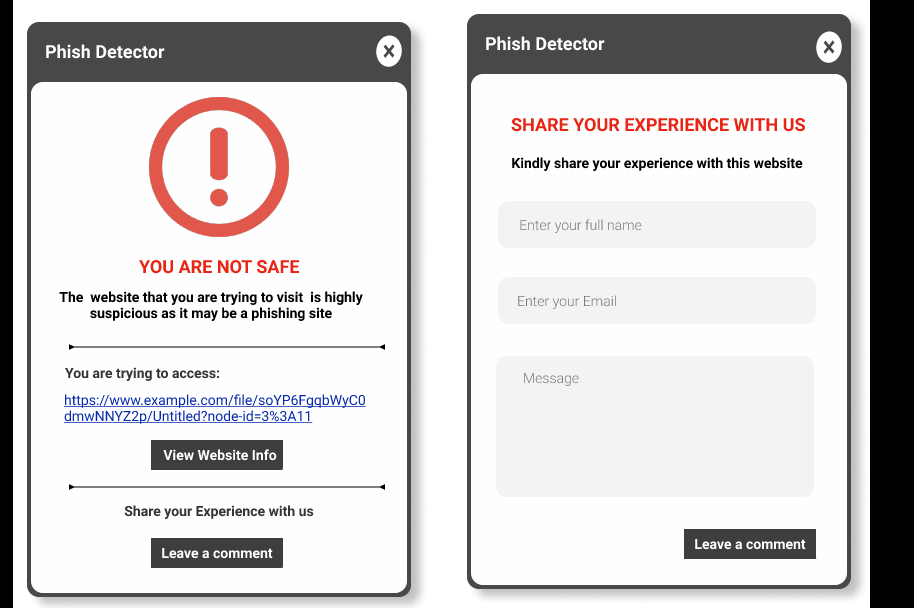
[===|-------|===]

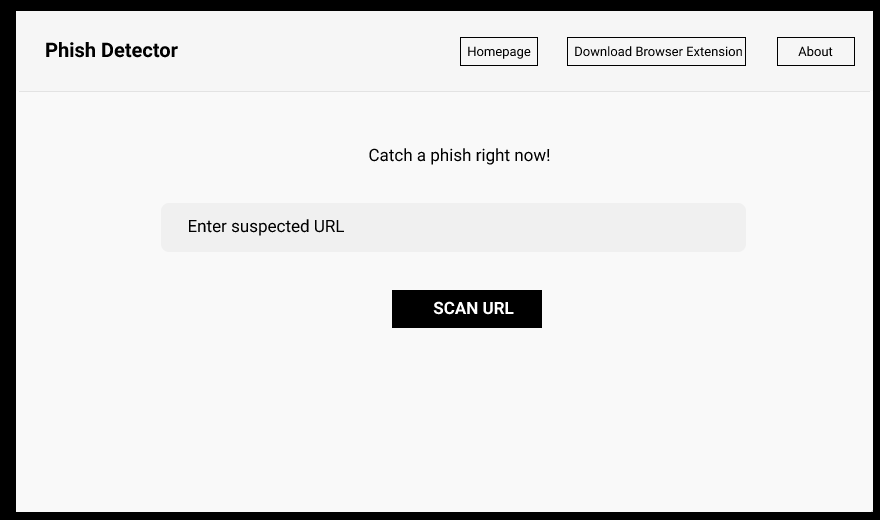
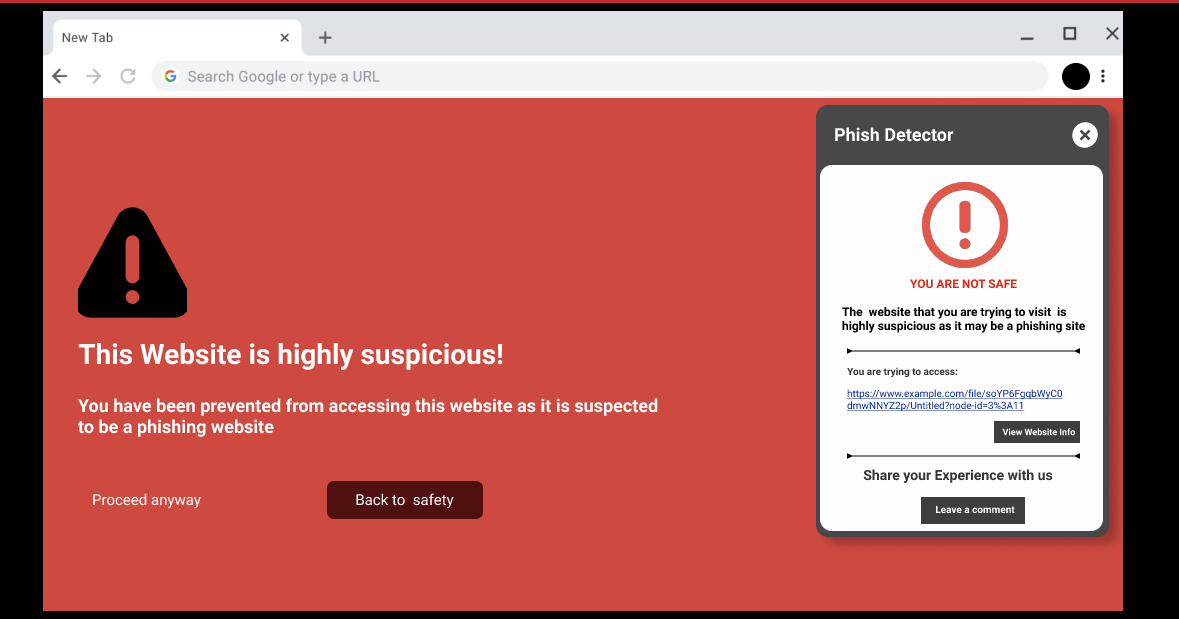
5.3. Database Design



5.4. User Interface Design (UI design of major system components and forms if any)

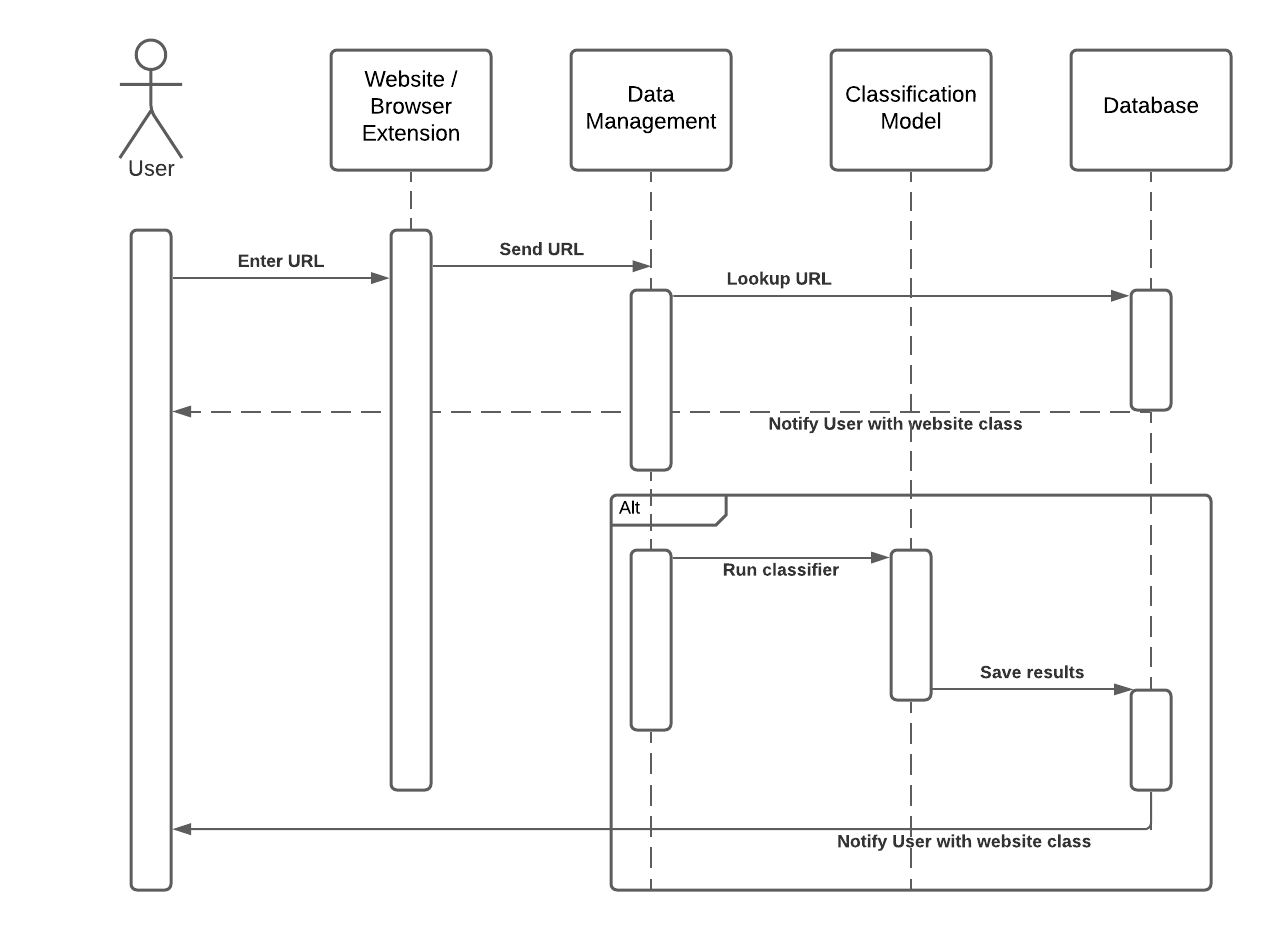






5.5. Other diagrams that help explain the system

SEQUENCE DIAGRAM



ACTIVITY DIAGRAM

