Chapter 4

Analysis Modeling

4.1 Functional Modeling

Data Flow Diagram

Data flow diagrams provide a graphical representation of how information moves between processes in a system. A Data Flow Diagram shows what kind of information will be input to and output from the system, where the data will come from and go to, and where the data will be stored.

Level 0 DFD for detecting phishing websites using data mining

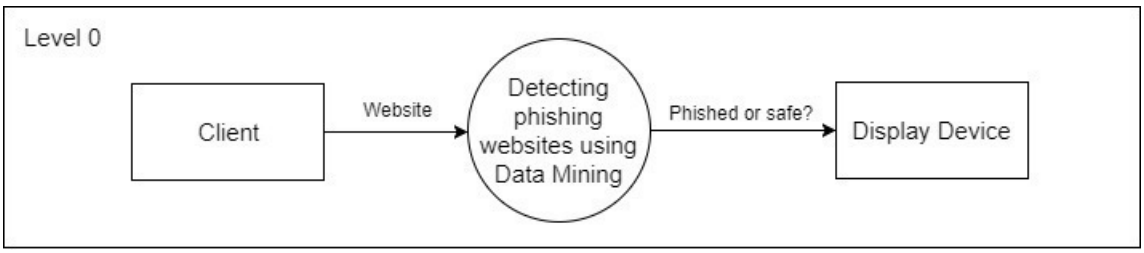


Figure 4.1.1: DFD Level 0

In this level 0 data flow diagram, the whole system is represented with the help of input, processing and output. The input to system is the URL entered by the end user. The system should display appropriate alert on the web browser if a phished website is detected.

Level 1 DFD for detecting phishing websites using data mining

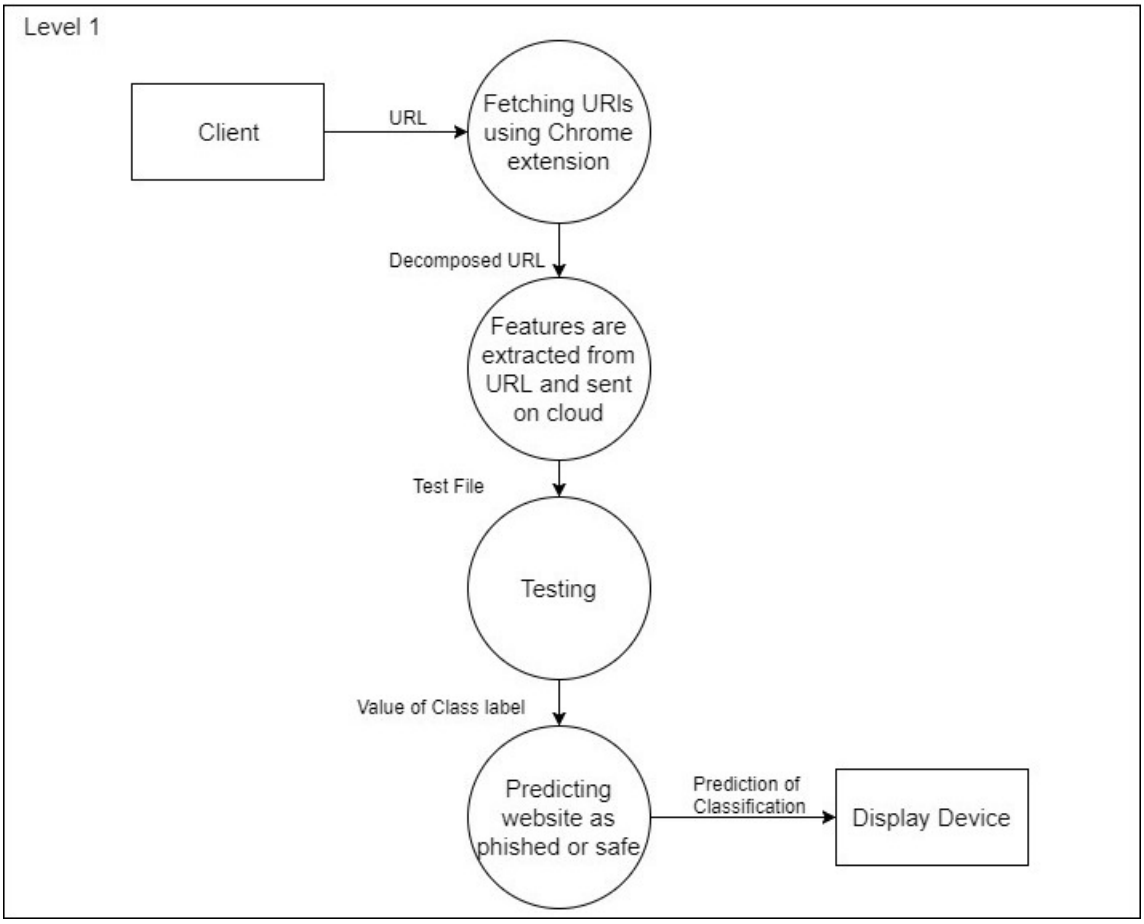


Figure 4.1.2: DFD Level 1

In level 1 data flow diagram, we have shown the process for phishing detection. When the client visits the URL, the Chrome extension will fetch the URL. The URL attributes are then extracted from the URL by the Chrome extension. These features are used as test data. Based on this, classifier predicts the website as safe or phished.

Level 2 DFD for detecting phishing websites using data mining

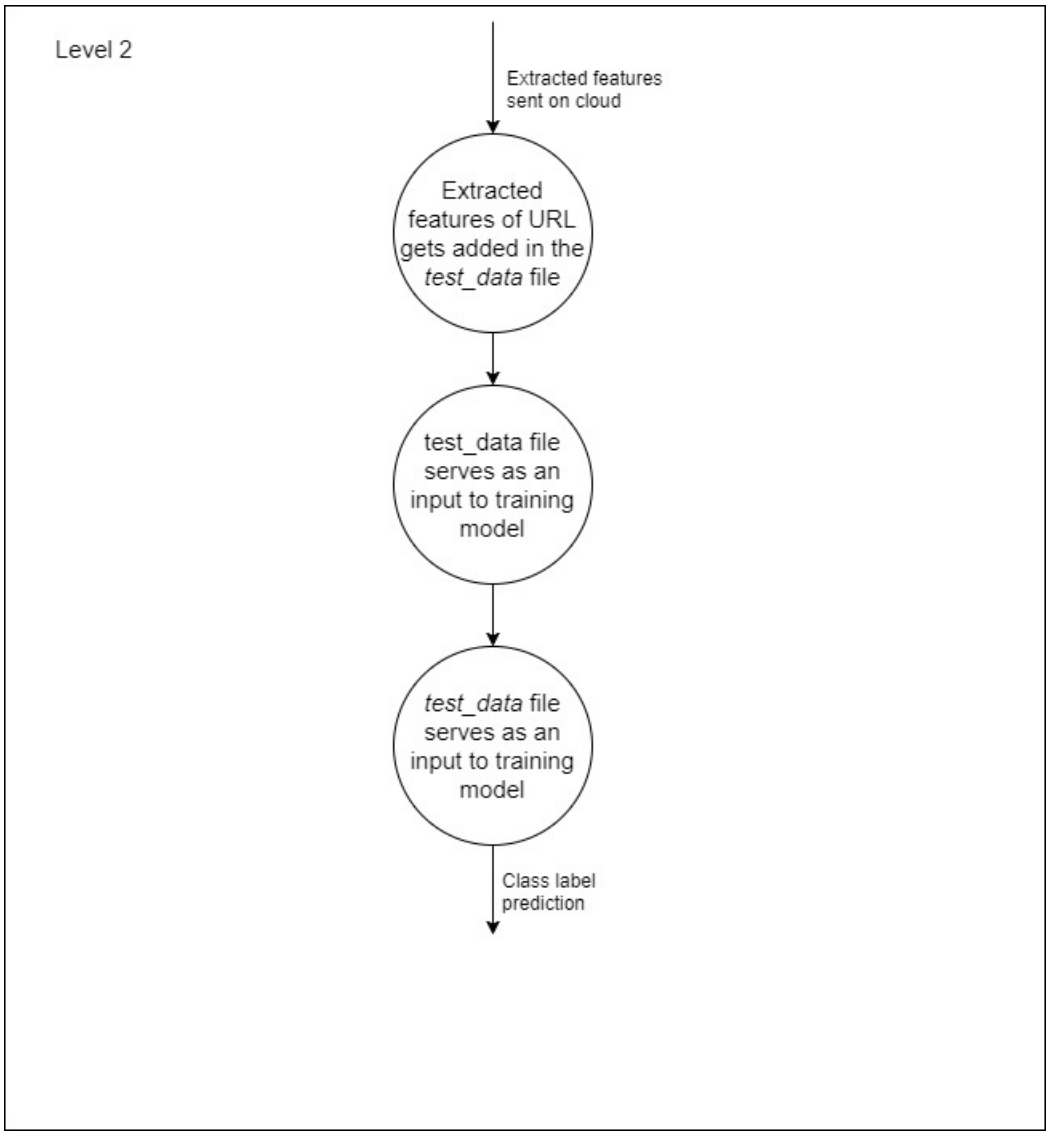


Figure 4.1.3: DFD Level 2

In level 2 data flow diagram, we have shown the detailed testing process. When the Chrome extension extracts the URL features, a test data file is created where the extracted URL attributes are added. (Each time a user visits a URL, a new *test data* file will be created which will only contain the visited URL as a *test data*). This file is then sent on the cloud for testing. The *test data* file acts as an input to the trained model. The model will then predict the value of the class label, which predicts the website as safe or phished. If the website is phished, the alert message is given to the user.

4.2 Activity Diagram:

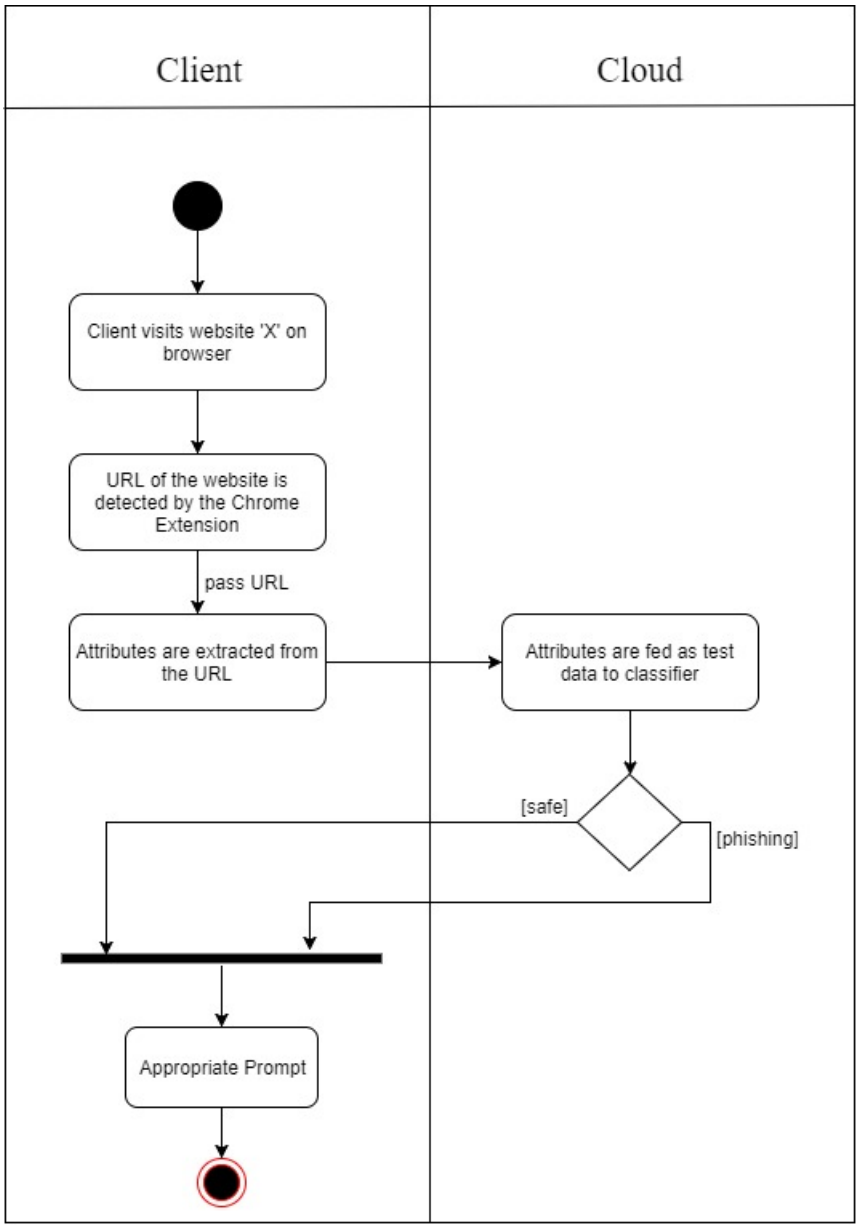


Figure 4.2.1: Activity Diagram

In figure 4.2, we have shown the activity diagram for Detecting Phishing websites using Data Mining. It is divided into client side and cloud side, thereby segregating the tasks for both. The client visits the website. The chrome extension on web browser fetches the URL and extracts the attributes from it. These attributes are sent on cloud where it is tested by the trained model residing on the cloud.

4.3 Flowchart:

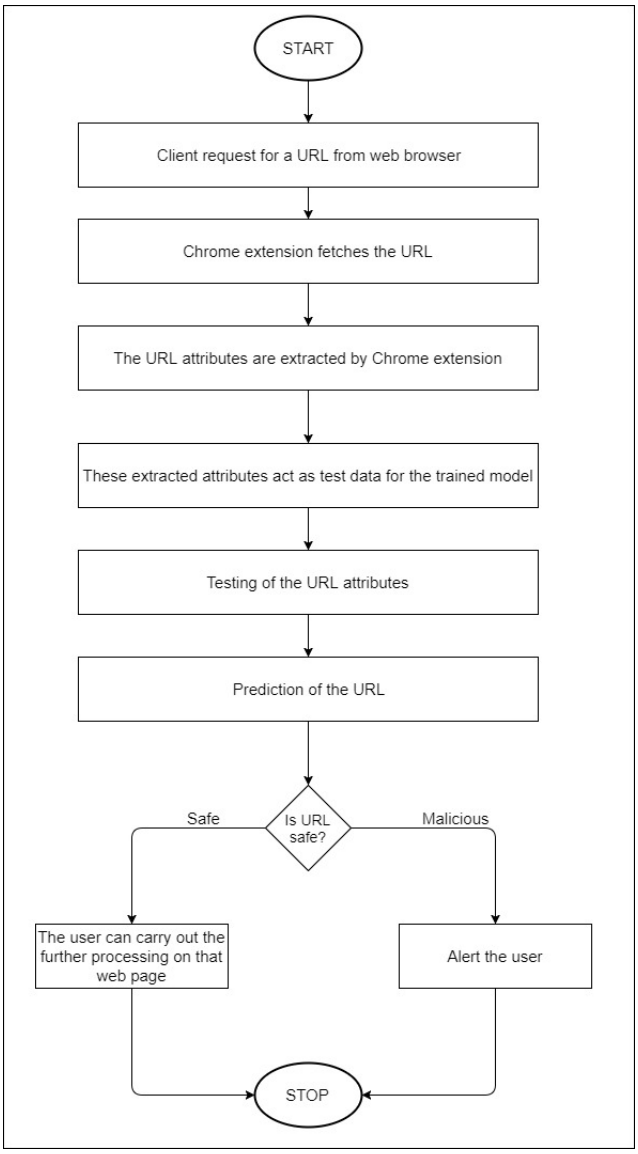
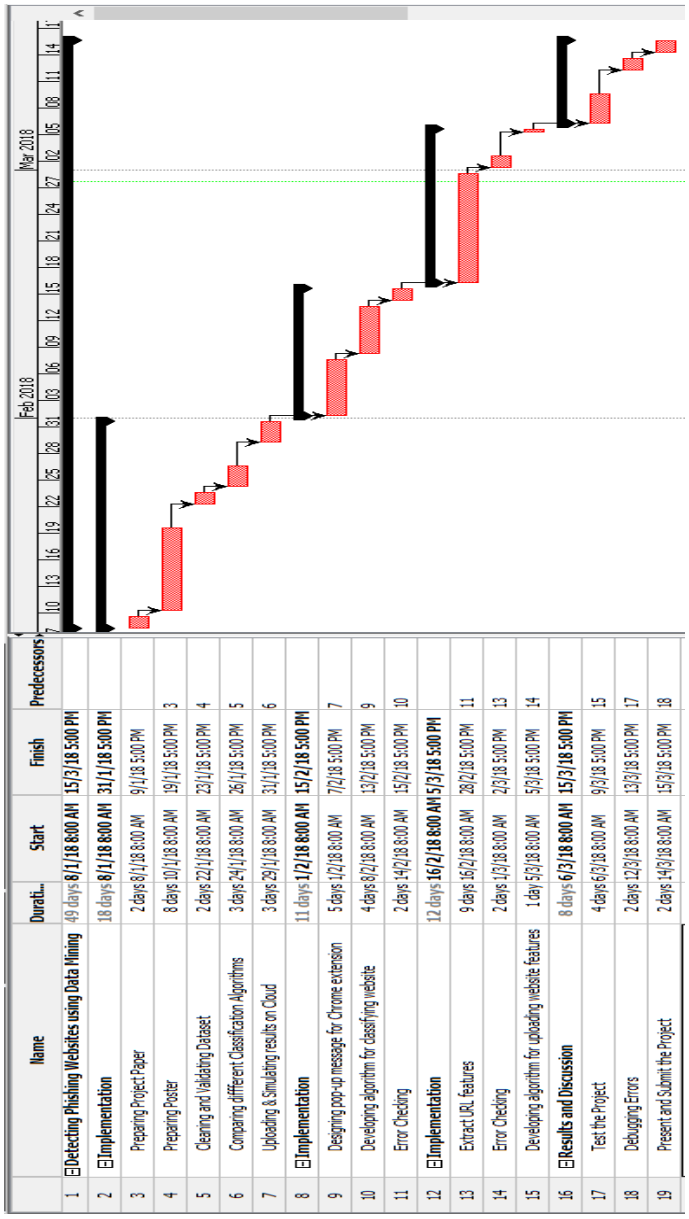


Figure 4.3.1: Flowchart for Phishing Detection using Data Mining

In figure 4.3.1, we have shown the complete flow of the project. To start with, the client visits a URL. As soon as the client visits the URL, the chrome extension on users browser will fetch that URL and extract the attributes of the URL. These attributes are then sent on the cloud where the classifier is deployed. The trained model then performs the testing of these URL attributes and predicts the legitimacy of the URL. It then alerts the user regarding the same. It is up to the user to proceed even if we alert about the phished website.

4.4 Timeline Chart:



	Name	Durati...	Start	Finish	Predecessors
1	Detecting Phishing Websites using Data Mining	49 days	8/1/18 8:00 AM	15/3/18 5:00 PM	
2	Implementation	18 days	8/1/18 8:00 AM	31/1/18 5:00 PM	
3	Preparing Project Paper	2 days	8/1/18 8:00 AM	9/1/18 5:00 PM	
4	Preparing Poster	8 days	10/1/18 8:00 AM	19/1/18 5:00 PM	3
5	Clearing and Validating Dataset	2 days	22/1/18 8:00 AM	23/1/18 5:00 PM	4
6	Comparing different Classification Algorithms	3 days	24/1/18 8:00 AM	26/1/18 5:00 PM	5
7	Uploading & Simulating results on Cloud	3 days	29/1/18 8:00 AM	31/1/18 5:00 PM	6
8	Implementation	11 days	1/2/18 8:00 AM	15/2/18 5:00 PM	
9	Designing pop-up message for Chrome extenson	5 days	1/2/18 8:00 AM	7/2/18 5:00 PM	7
10	Developing algorithm for classifying website	4 days	8/2/18 8:00 AM	13/2/18 5:00 PM	9
11	Error Checking	2 days	14/2/18 8:00 AM	15/2/18 5:00 PM	10
12	Implementation	12 days	16/2/18 8:00 AM	5/3/18 5:00 PM	
13	Extract URL features	9 days	16/2/18 8:00 AM	28/2/18 5:00 PM	11
14	Error Checking	2 days	1/3/18 8:00 AM	2/3/18 5:00 PM	13
15	Developing algorithm for uploading website features	1 day	5/3/18 8:00 AM	5/3/18 5:00 PM	14
16	Results and Discussion	8 days	6/3/18 8:00 AM	15/3/18 5:00 PM	
17	Test the Project	4 days	6/3/18 8:00 AM	9/3/18 5:00 PM	15
18	Debugging Errors	2 days	12/3/18 8:00 AM	13/3/18 5:00 PM	17
19	Present and Submit the Project	2 days	14/3/18 8:00 AM	15/3/18 5:00 PM	18

Figure 4.4.1: Timeline chart

Good planning spreads the necessary work over a reasonable period of time and allows everyone to work at a consistent, sustainable pace. A Timeline is a clean and concise visual representation of a series of events. It helps to arrange large chunks of time and see the overall plan easily. A timeline is typically divided into chunks, each ending with a milestone.

Here we have shown the Timeline chart, which shows the basic schedule of our project. The project has been divided into the phases and sub phases. The corresponding Gantt chart is also shown which gives an idea as to which phase has to be given more time.