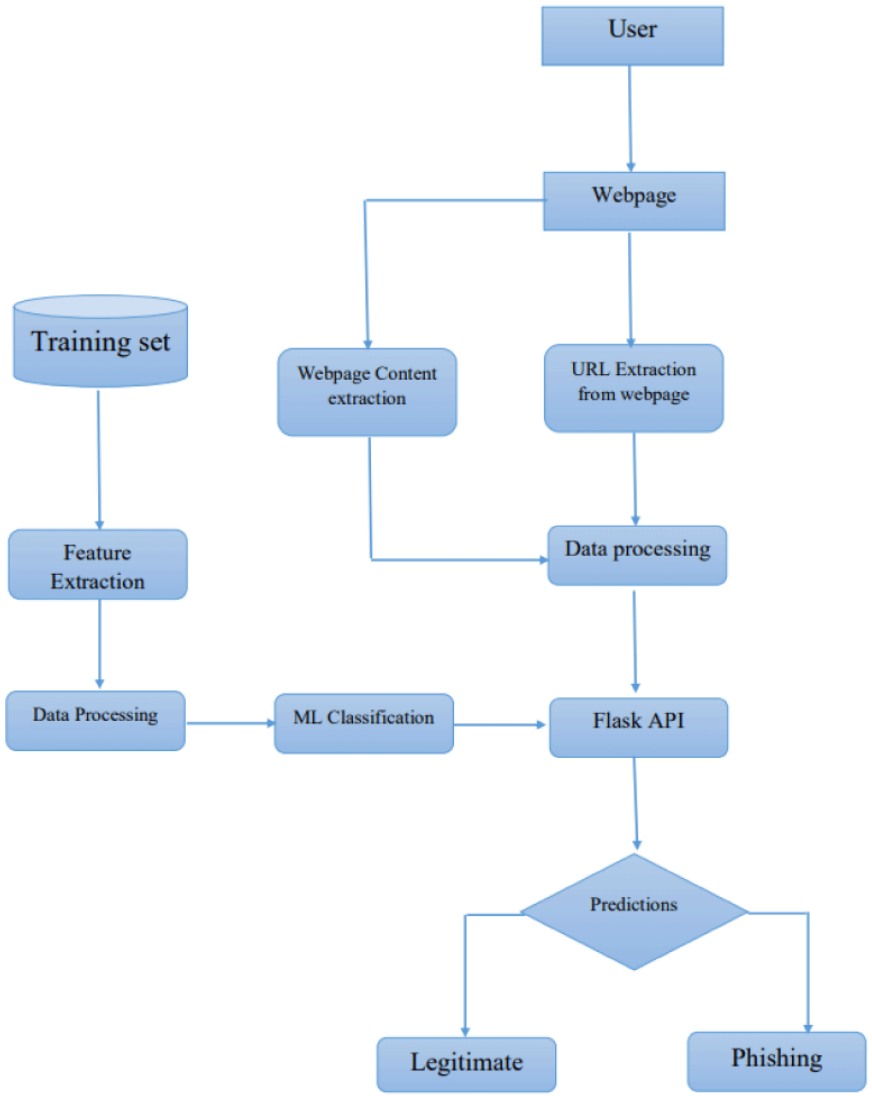
Project Design Phase-II
Data Flow Diagram & User Stories

Date	04 NOVEMBER 2022
Team ID	PNT2022TMID27211
Project Name	WEB PHISING DETECTION
Maximum Marks	4 Marks

Data Flow Diagrams:

The technique comprises of host based, page based and lexical feature extraction of collected websites. The primary step is the collection of phishing and benign websites. In the host-based approach, admiration based and lexical based attributes extractions are performed to form a database of attribute value. This database consists of knowledge mined that uses different machine learning techniques. On evaluating the algorithms, a selective classifier is opted and is implemented in Python Phishing is a major problem, which uses both social engineering and technical deception to get users' important information such as financial data, emails, and other private information. Phishing exploits human vulnerabilities; therefore, most protection protocols cannot prevent the whole phishing attacks. Many of them use the blacklist/whitelist approach, however, this cannot detect zero-hour phishing attacks, and they are not able to detect new types of phishing attacks. Data mining techniques have provided outstanding performance in many applications, e.g., data security and privacy, game theory, blockchain systems, healthcare, etc. Due to the recent development of phishing detection methods, various machine learning-based techniques have also been employed to investigate the legality of websites.


User
Webpage
Training set
Webpage Content extraction
URL Extraction from webpage
Feature Extraction
Data processing
Data Processing
ML Classification
Flask API
Predictions
Legitimate
Phishing

User Stories

Use the below template to list all the user stories for the product

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through Facebook	I can register & access the dashboard with Facebook Login	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail		Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password		High	Sprint-1
	Dashboard					
Customer (Web user)	User input	USN-1	As a user, I can enter the required URL in the box while awaiting validation	I can access the website without any problem	High	Sprint-1
Customer Care Executive	Feature extraction	USN-1	In the event that nothing is discovered during comparison we can extract features using a heuristic and a visual similarity technique	As a user I can have comparison between websites for security	High	Sprint-1
Administrator	prediction	USN-1	The model will use machine learning algorithms like a logistics regression and KNN to forecast the urls of the website	I can accurately forecast the specific algorithms in this way	High	Sprint-1
	classifier	USN-2	To create the final product, i will now feed all of the model output to classifier	I'll use this to identify the appropriate classifier for generating the outcome	Medium	Sprint-2