Project Design Phase-II
Technology Stack (Architecture & Stack)

Date	29 October 2022
Team ID	PNT2022TMID49605
Project Name	Web phishing detection
Maximum Marks	4 Marks

Technical Architecture:

The Deliverable shall include the architectural diagram as below

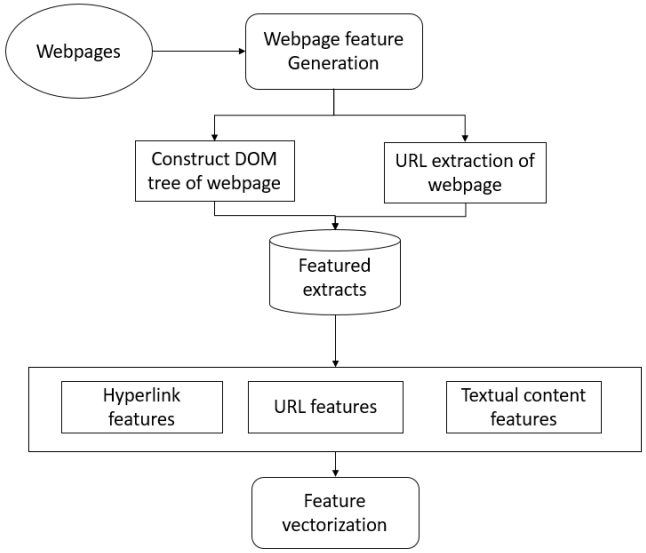

Fig 1: PREPROCESSING PHASE

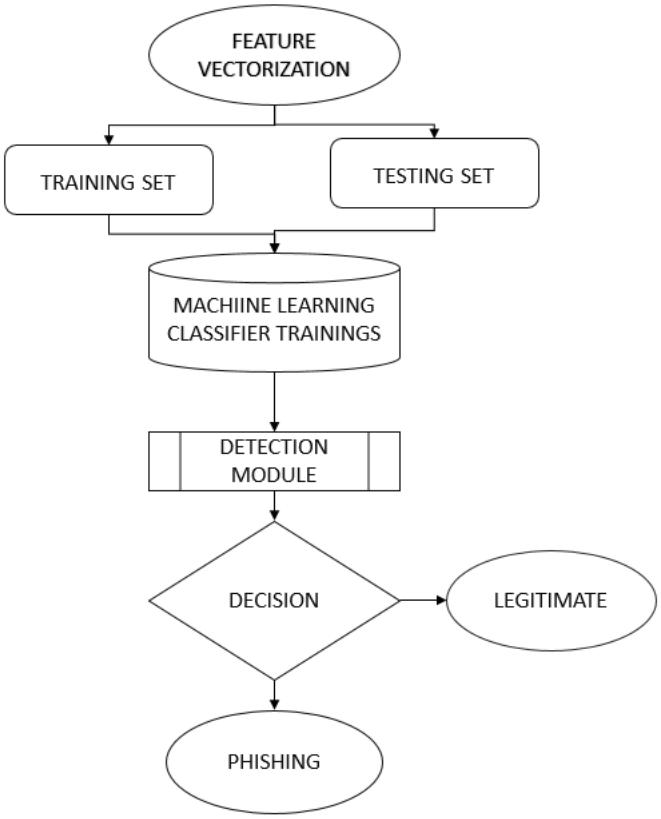

FIG 2: DETECTION PHASE

Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1.	User Interface	How user interacts with application e.g. Web UI, Mobile App, Chatbot etc.	HTML, CSS, JavaScript, Bootstrap, React.JS .
2.	Application Logic-1	Logic for a process in the application	Java / Python
3.	Application Logic-2	Logic for a process in the application	IBM cloud, Python flask
4.	Application Logic-3	Logic for a process in the application	IBM Watson Assistant
5.	Database	Data Type, Configurations etc.	MySQL
6.	Cloud Database	Database Service on Cloud	IBM DB2, IBM Cloud
7.	File Storage	File storage requirements	Local Filesystem
8.	External API-1	Purpose of External API used in the application	Google API, Apple REST API etc.
9.	Machine Learning Model	Purpose of Machine Learning Model	Gradient descent, Random Forest, Logistic Regression, Naive Bayes, and Ensemble of Random Forest and SVM classifiers.
10.	Infrastructure (Server / Cloud)	Application Deployment on Local System / Cloud	Kubernetes

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	List the open-source frameworks used	Zphisher and gophish
2.	Security Implementations	List all the security / access controls implemented, use of firewalls etc.	e.g. SHA-256, Encryptions, IAM Controls, OWASP .
3.	Scalable Architecture	Justify the scalability of architecture (3 – tier, Micro-services)	Flask API, python, mysql are the 3-tier micro-service.
4.	Availability	Justify the availability of application (e.g. use of load balancers, distributed servers etc.)	IBM cloud load balancer.
5.	Performance	Design consideration for the performance of the application (number of requests per sec, use of Cache, use of CDN's) etc.	Blacklists, visual similarity ,machine learning techniques.