Project Design Phase-I
Proposed Solution Template

Date	30 October 2022
Team ID	PNT2022TMID20856
Project Name	Project – Web Phishing Detection
Maximum Marks	2 Marks

Proposed Solution Template:

Project team shall fill the following information in proposed solution template.

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Phishing is a major issue that employs both social engineering and technical deception to obtain sensitive information such as financial data, emails, and other private information form users. In order to detect those phishing websites, we proposed an intelligent, flexible and effective system that is based on using classification algorithms. Our system will detect whether the site is true website or not.
2.	Idea / Solution description	The algorithms can be chosen according to the objective. As the dataset which we are using is a Classification dataset, we can use the following algorithms, Logistic Regression, Random Forest Regression / Classification, Decision Tree Regression / Classification, K-Nearest Neighbours, Support Vector Machine In order to get appropriate predictions, the dataset can be trained with any of the above algorithms. Then the model is built using the above any one of the algorithms. After the model is built, we will be integrating it to a web application so that normal users can also use it to know if any website is phishing or safe in a no-code manner.
3.	Novelty / Uniqueness	In the flask application, the URL is taken from the HTML page and it is scraped to get the different factors or the behaviour of the URL. These factors are then given to the model to know if the URL is phishing or safe and is sent back to the HTML page to notify the user. When the URL is given,

		the model analyses and gives the output whether it is a phishing or legitimate website.
4.	Social Impact / Customer Satisfaction	Based on the customer's requirements the model will be built to predict the phishing websites using Machine Learning Algorithms. Our application will provide the users to classify the phishing and legitimate sites and will alert them if the site is malicious.
5.	Business Model (Revenue Model)	A number of methods for detecting and filtering phishing attacks have been proposed in the literature. Researchers are still working on for a solution to protect customers from phishing attacks and produce better results. It may be easier to detect phishing websites if we're able to identify the specific patterns and characteristics they depict. Machine learning approaches can be used to solve the classification problem of identifying such aspects.
6.	Scalability of the Solution	We proposed an intelligent, flexible and effective system that is based on using classification algorithms. We implemented classification algorithms and techniques to extract the phishing datasets criteria to classify their legitimacy. The phishing website can be detected based on some important characteristics like URL and domain identity, and security and encryption criteria in the final phishing detection rate. Once a user makes a transaction online when he makes payment through an website our system will use a data mining algorithm to detect whether the website is a phishing website or not.