Project Design Phase-I
Proposed Solution Template

Date	24 September 2022
Team ID	PNT2022TMID34894
Project Name	Project - Web Phishing Detection
Maximum Marks	2 Marks

Proposed Solution Template:

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

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Hackers install malicious software on computers to steal credentials, often using systems to intercept username and passwords of consumers' online accounts. Phishers use multiple methods, including email, Uniform Resource Locators (URL), instant messages, forum postings, telephone calls, and text messages to steal user information. The structure of phishing content is similar to the original content and trick users to access the content in order to obtain their sensitive data. The primary objective of phishing is to gain certain personal information for financial gain or use of identity theft. Phishing attacks are causing severe economic damage around the world. Moreover, Most phishing attacks target financial/payment institutions and webmail, according to the Anti-Phishing Working Group (APWG) latest Phishing pattern studies .
2.	Idea / Solution description	Malicious URLs on the Internet can be easily identified by analysing it through Machine Learning (ML) technique . The conventional URL detection approach is based on a blacklist (set of malicious URLs) obtained by user reports or manual opinions. On the one hand, the blacklist is used to verify an URL and on the other hand the URL in the blacklist is updated frequently. However, the numbers of malicious URLs not on the blacklist are increasing significantly. For instance, cybercriminals can use a Domain Generation Algorithm (DGA) to circumvent the blacklist by

		creating new malicious URLs. Thus, an exhaustive blacklist of malicious URLs is almost impossible to identify the malicious URLs. Thusnew malicious URLs cannot be identified with the existing approaches. Researchers suggested methods based on the learning of computers to identify malicious URLs to resolve the limitations of the system based on the blacklist . Malicious URL detection is considered a binary classification task with two-class predictions: malicious and benign. The training of the ML method consists of finding the best mapping between the d-dimensional vector space and the output variable . This strategy has a strong generalisation capacity to find unknown malicious URLs compared to the blacklist approach.Recurrent Neural Network (RNN)—Long Short-Term Memory (LSTM) is one of the ML techniques that presents a solution for the complex real—time problems . LSTM allows RNN to store inputs for a larger period . It is similar to the concept of storage in computer. In addition, each feature will be processed according to the uniform distribution . The combination of RNN and LSTM enables us to extract a lot of information from a minimum set of data. Therefore, it supports a phishing detection system to identify a malicious site in a shorter duration.In comparison to most previous approaches, researchers focus on identifying malicious URLs from the massive set of URLs. Therefore, we propose a Recurrent Neural Network (RNN) based URL detection approach.
3.	Novelty / Uniqueness	To develop a novel approach to detect malicious URLs and alert users.We are applying ML techniques in the proposed approach in order to analyse the real time URLs and produce effective results. To implement the concept of RNN, which is a familiar ML technique that has the capability to handle huge amount of data.
4.	Social Impact / Customer Satisfaction	Phishing is one of the top cyber-crimes that impact consumers and businesses all around the world. It is the most common scam on the Internet. Phishing is known as the

		process in which someone attempts to obtain sensitive information such as usernames, passwords, social security number or financial information and personal information such as birthdates, name and addresses by masking themselves as a trustworthy or familiar entity. With social networking on the rise, people are sharing their personal information everywhere, and have no idea if a website is truly what it seems to be. Phishing scam is an important topic because it is no longer something that can only be done by hackers, but by anyone with internet access. A successful phishing attack can have disastrous consequences for the victims leading to financial losses and identity theft. In this report, it will be going to cover the use of Phishing Scams in the society and the potential impacts of the Phishing Scams in both current and future.
5.	Business Model (Revenue Model)	Timely information about the data's hacked and amount of information taken from a user and thereby minimises the costs of creating the website. Monitoring of the potential risk areas and an early detection of the hackers can significantly shorten the risk of hacking. It may not look like it at first with the initial upfront investment that can be substantial, however, over the long run a well maintained and tested commercial web pishing system can lower the insurance cost thus lowering overhead cost. With a high quality, well installed and well maintained web pishing system this minimises or even erases the unnecessary interruption of alarm. This reduces the amount of damage to the property.
6.	Scalability of the Solution	This solution is scalable enough to fit the Security issues by constructing the best website. The cost of establishing the website and maintaining all the programs may be high . It is acceptable to fit them over any place and any resources.