Project Initialization and Planning Phase

Date	23 September 2024
Team ID	LTVIP2024TMID24973
Project Title	Detection of Phishing Websites from URLs Using Machine learning
Maximum Marks	3 Marks

Project Proposal (Proposed Solution) template

This project proposal outlines a solution to address a specific problem. With a clear objective, defined scope, and a concise problem statement, the proposed solution details the approach, key features, and resource requirements, including hardware, software, and personnel.

Project Overview	
Objective	To develop a machine learning-based solution that automatically detects and flags phishing websites by analyzing URLs.
Scope	This project will focus on collecting a dataset of URLs, developing and training machine learning models, and creating a user interface for users to input URLs for phishing detection. The project will be completed over a six-month period.
Problem Statement	
Description	Phishing websites impersonate legitimate sites to steal sensitive information from users. As these attacks become more sophisticated, traditional detection methods are often inadequate, leading to increased risk for users and organizations.
Impact	Implementing an effective detection system will help users avoid phishing attempts, enhancing online safety and reducing the incidence of identity theft and financial loss.
Proposed Solution	
Approach	□ **Data Collection**: Gather a dataset of known phishing and legitimate URLs from various sources. □ **Feature Extraction**: Analyze URLs to extract relevant features (e.g., length, use of HTTPS, presence of suspicious characters).

	□ **Model Training**: Use supervised learning techniques to train models on the dataset for classification. □ **Implementation**: Develop a web-based interface for users to input URLs and receive phishing risk assessments.
Key Features	□ **Real-time URL Analysis**: Immediate classification of input URLs as phishing or legitimate. □ **Feature Visualization**: Show users how certain features contribute to the phishing risk assessment. □ **User-Friendly Interface**: Simple design for easy URL input and display of results. □ **API Integration**: Allow other applications to utilize the phishing detection service.

Resource Requirements

Resource Type	**Description**	**Specification/Allocation**
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	e.g., 2 x NVIDIA V100 GPUs
Memory	RAM specifications	e.g., 16 GB
Storage	Disk space for data, models, and logs	e.g., 1 TB SSD
Software		
Frameworks	Python frameworks	e.g., Flask,
Libraries	Additional libraries	e.g., scikit-learn, pandas, numpy
Development Environment	IDE, version control	e.g., Jupyter Notebook, Git
Data		
Data	Source, size, format	Public datasets (e.g., Phish Tank), 100,000 URLs

This project proposal aims to create a robust solution for detecting phishing websites through URL analysis, significantly enhancing online security for users.