

WEB PHISHING DETECTION

Team ID	PNT2022TMID33764
Project Name	Project – Web Phishing Detection

PRE-REQUISITES :

Anaconda Navigator :

Anaconda Navigator is a free and open-source distribution of the Python and R programming languages for data science and machine learning related applications. It can be installed on Windows, Linux, and MAC OS. Conda is an open-source, cross-platform, package management system. Anaconda comes with great tools like JupyterLab, Jupyter Notebook, QtConsole, Spyder, Glueviz, Orange, Rstudio, Visual Studio Code. For this project, we will be using Jupyter notebook and Spyder To install Anaconda navigator and to know how to use Jupyter Notebook & Spyder using Anaconda watch the video.

Jupyter Notebook:

Jupyter lab is a project with goals to develop open-source software, open standards, and services for interactive computing across multiple programming languages. It is a web-based interactive development environment for notebooks, code, and data. Its flexible interface allows users to configure and arrange workflows in data science, scientific computing, computational journalism, and machine learning.

THE FOLLOWING PACKAGES ARE REQUIRED TO BUILD MACHINE LEARNING MODELS:

Sklearn:

Scikit-learn (formerly scikits.learn and also known as sklearn) is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support-vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

Numpy:

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

Pandas:

Pandas is a software library written for the Python programming language for data manipulation and analysis. It is most widely used for data science/data analysis and machine learning tasks. In particular, it offers data structures and operations for manipulating numerical tables and time series.

Matplotlib:

Matplotlib is a visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack. It consists of several plots like line, bar, scatter, histogram etc.

Flask:

Flask is a micro web framework written in Python. It is classified as a micro framework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where preexisting third-party libraries provide common functions. However, Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools.

Google Colab:

Google Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. It allows you to combine executable code and rich text in a single document, along with images, HTML, and more.

VS Code:

Visual Studio Code, also commonly referred to as VS Code, is a source-code editor made by Microsoft with the Electron Framework, for Windows, Linux and MAC OS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded git. Users can change the theme, keyboard shortcuts, preferences, and install extensions that add additional functionality.