WEB PHISHING DETECTION

Team ID	PNT2022TMID33764
Project Name	Project – Web Phishing Detection

PRIOR KNOWLEDGE :

Supervised Learning :

Supervised learning is the type of machine learning in which machines are trained using well "labeled" training data, and on the basis of that data, machines predict the output. The labeled data means some input data is already tagged with the correct output.

In supervised learning, the training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly. It applies the same concept as a student learns in the supervision of the teacher.

Supervised learning is a process of providing input data as well as correct output data to the machine learning model. The aim of a supervised learning algorithm is to find a mapping function to map the input variable(x) with the output variable(y).

In the real-world, supervised learning can be used for Risk Assessment, Image classification, Fraud Detection, spam filtering, etc.

Types of supervised Machine learning Algorithms:

Supervised learning can be further divided into two types of problems:

➢ Regression

➢ Classification

1. Regression :

Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc. Below are some popular Regression algorithms which come under supervised learning

➢ Linear Regression

➢ Regression Trees

> ➢ Non-Linear Regression

> ➢ Bayesian Linear Regression

> ➢ Polynomial Regression

2. Classification :

Classification algorithms are used when the output variable is categorical, which means there are two classes such as Yes-No, Male-Female, True-false, etc.

Spam Filtering,

> ➢ Random Forest

> ➢ Decision Trees

> ➢ Logistic Regression

> ➢ Support vector Machines

Clustering :

It is basically a type of unsupervised learning method. An unsupervised learning method is a method in which we draw references from datasets consisting of input data without labeled responses. Generally, it is used as a process to find meaningful structure, explanatory underlying processes, generative features, and groupings inherent in a set of examples.

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them.

Logistic Regression :

Logistic regression is one of the most popular machine learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.

Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, True or

False, etc. But instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.

Unsupervised Learning :

Unsupervised Learning is the training of a machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Here the task of the machine is to group unsorted information according to similarities, patterns, and differences without any prior training of data.

Unlike supervised learning, no teacher is provided that means no training will be given to the machine. Therefore the machine is restricted to finding the hidden structure in unlabeled data by itself.

FLASK Library :

A Flask extension typically has flask in its name as a prefix or suffix. If it wraps another library, it should include the library name as well. This makes it easy to search for extensions, and makes their purpose clearer.

A general Python packaging recommendation is that the install name from the package index and the name used in import statements should be related. The import name is lowercase, with words separated by underscores (_). The install name is either lower case or title case, with words separated by dashes (-). If it wraps another library, prefer using the same case as that library's name.