Name: Smitesh Patil

Class: 1CSD

Student Id: 22223696

1. Have a look at the given data, understand the problem based on dependent variable and select a machine learning category that can solve the task/problem. Briefly explain why do you think it is the correct ML category for this problem?

**Answer**: Given data contains four independent variables, namely sepal length, sepal width, petal length, petal width. These variables determine value of our dependent variable “target/label”. Target/label contains two values setosa and virginica. Theses are two types of flowers, and the dependent variables are features that determine the type of the flower. In the given dataset, we have the target label given; hence, it is a case of supervised learning problem. Moreover, the target label is a discrete variable as it contains two values setosa and virginica, hence it is a classification task.

1. Explore and report the data and its distribution among training and testing data. Can we call it imbalanced dataset, explain your answer (yes/no) briefly? [2 Mark]

**Answer:** Datasets that contain uneven distribution of target class variable are called imbalanced dataset.

Graphical user interface, text, application

Description automatically generated

In both the datasets, the target class variables have a even distribution of setosa and virginica observations. Thus the dataset is not imbalanced.