**Assignment Problems: Day 14**

The following problems require access to the file named ‘Dataset\_Day14.csv’, provided with this assignment.

This is the Iris dataset with150 data points (same dataset as Day 13)

It includes three iris species with 50 samples each as well as some properties about each flower. One flower species is linearly separable from the other two, but the other two are not linearly separable from each other.

The columns in this dataset are:

Id

SepalLengthCm

SepalWidthCm

PetalLengthCm

PetalWidthCm

Species

Problems to solve –

1. Treat outliers (optional for today) and check missing values if present and scale the data.
2. Fit the DBSCAN clusters for the default parameter values and also show the *Species* distribution in each of the default clusters. Default: (eps = 0.8 and min\_sample = 4)
3. Use nearest neighbour algorithm to find the most optimal value of ‘eps’ parameter.
4. Use the ‘eps’ value in (2.) and find the most optimal value of ‘min\_samples’ using silhouette score.
5. Find all the outliers using the DBSCAN algorithm.