# **BATCH -1**

# **Assignment - 1**

# **Species Segmentation with Cluster Analysis**

The Iris flower dataset is one of the most popular ones for machine learning. You can read a lot about it online and have probably already heard of it: <https://en.wikipedia.org/wiki/Iris_flower_data_set>

We didn't want to use it in the lectures, but believe that it would be very interesting for you to try it out (and maybe read about it on your own).

There are 4 features: sepal length, sepal width, petal length, and petal width.

Start by creating 2 clusters. Then standardize the data and try again. Does it make a difference?

Use the Elbow rule to determine how many clusters are there.

## Import the relevant libraries

## Load the data

Load data from the csv file: *'iris\_dataset.csv'*.

## Plot the data

For this exercise, try to cluster the iris flowers by the shape of their sepal.

*Use the 'sepal\_length' and 'sepal\_width' variables.*

# Clustering (unscaled data)

Separate the original data into 2 clusters.

# Standardize the variables ( standard scaler) or you can robust scaler as well – try to find how to do that of your own. Robust scaler is optional.

Import and use function from sklearn to standardize the data.

# Clustering (scaled data)

## Take Advantage of the Elbow Method

### Calculate and display WCSS

### Find the optimal number of clusters

### Create Visualization

Submit your answers at : <https://forms.gle/2s7K55zWNVn16YU2A>