# Lab Assignment 10 : k-means

This assignment is based on analyzing clustering techniques. The objective of this assignment is to become familiar with **clustering algorithm** and evaluate or compare their performance using *Weka*.

1. Select the "**segment-test.arff**" dataset in Weka and run the Simple K-means algorithm on it using 2, 4, 8, and 16 clusters with 2 different distance functions: ***EuclideanDistance*** and ***ManhattanDistance*** respectively with a 44% percentage split (in total you would need to run it 8 times – 4 clusters with each distance function). You can keep the default values of the remaining parameters like maximum number of iterations, random seed, etc.

   1.1 How many clusters do you get?
   1.2 Visualize the clusters (graph) and take a screenshot of your results.

2. In the Explorer application, open the **cpu.arff** data file and do the following:

   2.1 Cluster the data (using the simple k-Means algorithm, with **k=3**) and report on the nature and composition of the extracted clusters.

3. Load the 'weather.arff' data set and click on the 'Cluster' tab. Choose the 'SimpleKMeans' classifier with the default options. Under the 'Cluster mode' choose '**Classes to clusters evaluation'** to evaluate the clustering performance using the class labels provided in the data file. Click the 'Start' button.
   3.1 Explain the purpose of the menu under the option 'Classes to clusters evaluation' in **one sentence**.
   3.2 What is the number of clusters?
   3.3 What is the number of iterations?
   3.4 What class label is assigned to Cluster 0?
   3.5 What is the percentage of correctly clustered instances ?