Assignment-ML

Consider the iris data set and perform clustering on it. The iris dataset is of 150 instances represented by 4 features. The last column in the iris data set represents the label information. Group the dataset in an unsupervised way i.e., without considering the label information. Perform the clustering algorithm using the following algorithm i) K-means, ii) Bottom-up Hierarchical Clustering (BUHC) using single linkage, and iii) Fuzzy C means (FCM) (non-overlapping). Provide the number of cluster K as K=3, 4 and 5. Use the Silhouette score and Adjusted Rank Index (ARI) for evaluating the generated clusters. Report the score obtained below in a tabular form.

	Silhouette Score			ARI		
	K-means	BUHC	FCM	K-means	BUHC	FCM
K=3						
K=4						
K=5						

Normalize the dataset before clustering i.e., replace each element (x) in the dataset with its Z-value such that Z=(x-m)/d where m is the mean of the dataset and d is the standard deviation.

Please find the iris dataset (iris.data) from the below link:

https://archive.ics.uci.edu/ml/machine-learning-databases/iris/

Or

https://github.com/dipanjyoti/BSE_Assignment.git