Lab Dimention Reduction + Clustering

"A friend in need is a friend indeed." (Dimension reduction and clustering)

PCA

1. Get the dataset from this address : https://github.com/AttilaDSA/IntilaqDSAcademy/blob/master/Clustering%20Labs/Dimension%20reduction%20%2B%20Clustering%20Lab/voice.csv

2. Decompose the dataset into 2 parts: train set and test set.

3. Create two data frames from the train set : a df containing the voices treatment values without the label (male or female) and the second containing the labels.

4. Reduce the first dataframe’s dimensions while conserving 90% of its cumulative explained variance.

Clustering

1. Which one of these algorithms seems to be more adapted to cluster the voices into male voices and female voices?
   a. K-means
   b. Agglomerative
   c. DBSCAN

2. Use the appropriate algorithm to cluster the data then using the label dataframe, compute the success rate of the resulting clusters.

Logistic regression

1. Using logistic regression, train your model with the clustered values then test with the test set. How’s the accuracy of the classification?

2. Logistic regression can be used without clustering (since we already have the clusters at the beginning). So use a logistic regression algorithm to classify the test set after training it with the train set (without any pca or clustering). How’s th accuracy of this classification compared to the one before?

   NB : You can use **LogisticRegression** from **sklearn.linear_model**