

Lab 5

```
In [1]: import pandas as pd  
import numpy as np
```

In this lab, we will use the red wine quality dataset which is attached with the instructions, you can find the description at <https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009/data>.

Question 1

Make a data frame with first 11 columns from the original dataset You will apply the different clusters you learned in class to this dataset. Then, processing the dataset, you may deal with the null value, high related factors. (10)

```
In [ ]: # Write your code here
```

Question 2

Apply K-Means Cluster to the dataset, Apply K-Means Cluster to the dataset, print the Silhouette score (10).

```
In [ ]: # Write your code here
```

Question 3

Apply Hierarchical to the dataset, Apply K-Means Cluster to the dataset, print the Silhouette score.(10)

```
In [ ]: # Write your code here
```

Question 4

Apply DBScan to the dataset, print the Silhouette score and best cluster number. (10)

```
In [ ]: # Write your code here
```

Question 5

Now make cluster number = 10, redo the cluste process and print the Silhouette scores. (10)

```
In [ ]: # Write your code here
```

Question 6

Match the cluster results with the 12th column, "quality", to add a column named as "prediction". For example, find all the data with quality = 6, and count the number of cluster numbers if in total there are 120 data with quality = 6 and 100 of their cluster numbers are 2, then all the data with cluster number = 2 should be marked as 6 in the "prediction" column. (10)

In [5]: *# Write your code here*

Question 7

Now get the accarcy score with the "Prediction" column and "quality" column.(10)

In []: *# Write your code here*

Question 8

Now seporate the data with 80% training set and 20% testing set and train a ANN model with the training set and print the accuracy score of the testing set.(10)

In []: *# Write your code here*

Question 9

Compare the accuracy scores of the previous four models, which one is the best? (10)

Write your code here

Question 10

Do you think we can use the unsupervised cluster models to be a classifier in this way? Why?(10)

Write your code here