

Names :

First student	rowan hassan	2019482466
Second student	alaa Ossama	2019482712

We both shared every step of these project and did our best to seek the right answer as much as we could

Dataset link :

<https://archive.ics.uci.edu/ml/datasets/User+Knowledge+Modeling>

Dataset description:

It is a real dataset about the students' knowledge status about the subject of Electrical DC Machines so there are five factors that we which are used to rate the knowledge status

These five attributes are :

- 1)STG (The degree of study time for goal object materials)
- 2)SCG (The degree of repetition number of user for goal object materials)
- 3)STR (The degree of study time of user for related objects with goal object)

4)LPR (The exam performance of user for related objects with goal object)

5)PEG (The exam performance of user for goal objects)

Number of attributes: 5

The target:

Our desire is to rate the student knowledge status about the subject of Electrical DC Machines

Into 4 clusters

UNS (The knowledge level of user) (target value)

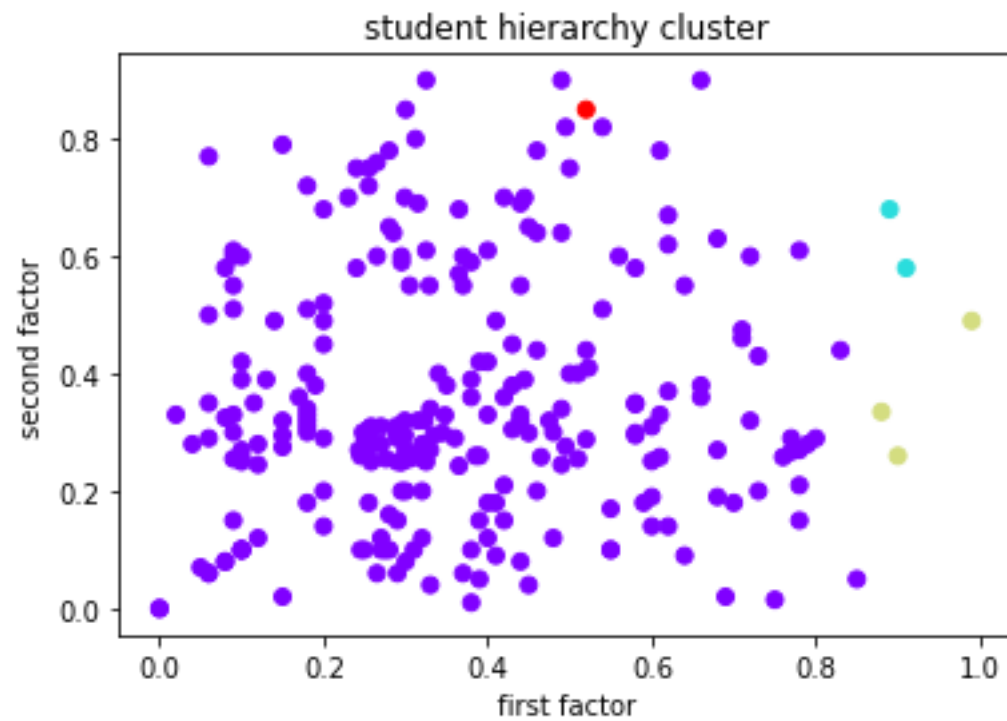
Very Low: 50

Low:129

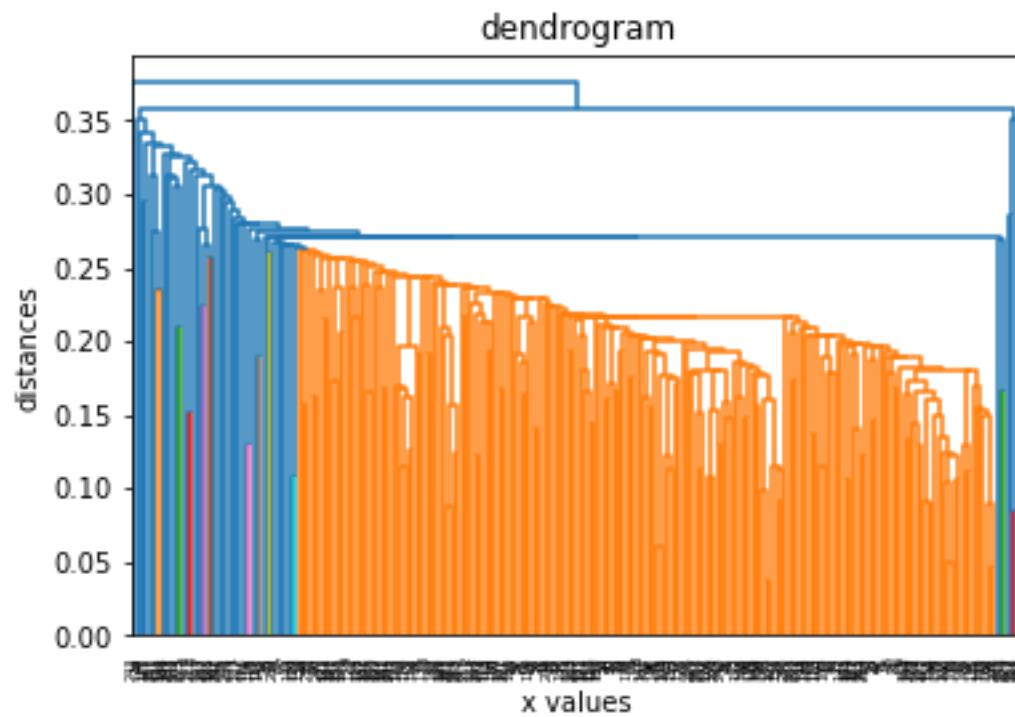
Middle: 122

High: 130

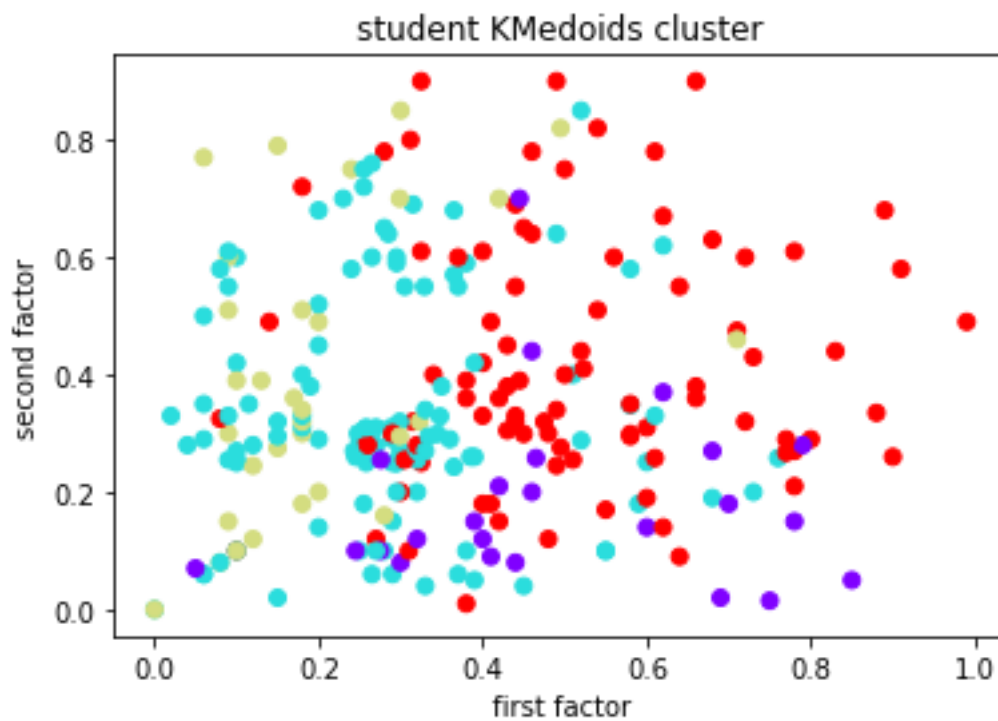
The plot of hierarchy



The plot of dendrogram



The plot of k-medoids



The explanation of our results:

hierarchal clustering is not easy to define levels as our data set so it's quite hard to recognize the clusters as it's not the best solution with outliers too so it is not the best visualization experience so we needed the assistance of the dendrograms which are great for visualizations ,And even hierarchical clustering wasn't the best solution for our data because the nature of leveled data but it was great for it's size as it's really fit big data

K-medoids clustering

does not scale well with large data so it wasn't the first solution for our data set but it works better with the smaller ones

and it is more robust to noises and outliers so it helped us to get our target value and medoid is the most centrally located object of the cluster with minimum sum of distances to other points and it was great for our multi leveled data

So; k-medoid is preferred with this data set

Comparison between hierarchal clustering graph and k-medoids graph:

<u>hierarchal clustering graph</u>	<u>k-medoids graph</u>
Not the best solution for the small data	Its Better in small data
Deal with big data	Cant deal with big data
Not good In having outliers	The best in having outliers
Difficult to define levels for clusters	Can be easy in defining levels for clusters
dendrograms are great for visualizations	Not have an assistant to help in visualization