Read Me:

To execute the Cluster for the records of GE

Step1: Find the ClusteringService module in the clustering module and execute using python.

CMD:

>>> python ClusteringServiceImplementationModarator.py

There are 3 (optional) arguments that you can pass

* **Cluster number:** number of cluster if you have any specific cluster set to be produced.

Example >>> python ClusteringServiceImplementationModarator.py ’20’ (for 20 clusters)

* **Stem:** If you prefer to have preprocessing done using the stem of the words

Example >>> python ClusteringServiceImplementationModarator.py ‘None’ ‘stem’(to stem)

* **Percentage of cluster based on category result:** If you already know the number of categories provided by the data analyst, then you can give the percentage on which the cluster will be created
  + 1. Example >>> python ClusteringServiceImplementationModarator.py ‘None’ ‘None’ ’60’
    2. If say there are 300 different categories provided by the data analyst and if you give 60% then (300/100)\*60 = 180 cluster will be formed.
    3. However, this cannot be ensured, if the number of documents/records are less to form so many cluster. In such case logic behind the scenes will override your percentage.

You can however check final percentage generated by the logic in the logs