

Prepare text data  
for Machine Learning  
with Scikit-Learn.

Apply Gaussian Mixture  
Models Clustering.

1.

2.

Apply K-Means  
clustering.

3.

1	15.26	14.84	0.871	5.763	3.312	2.221	5.22	1
2	14.88	14.57	0.8811	5.554	3.333	1.018	4.956	1
3	14.29	14.09	0.905	5.291	3.337	2.699	4.825	1
4	13.84	13.94	0.8955	5.324	3.379	2.259	4.805	1
5	16.14	14.99	0.9034	5.658	3.562	1.355	5.175	1
6	14.38	14.21	0.8951	5.386	3.312	2.462	4.956	1
7	14.69	14.49	0.8799	5.563	3.259	3.586	5.219	1
8	14.11	14.1	0.8911	5.42	3.302	2.7	5	1
9	16.63	15.46	0.8747	6.053	3.465	2.04	5.877	1
10	16.44	15.25	0.888	5.884	3.505	1.969	5.533	1
11	15.26	14.85	0.8696	5.714	3.242	4.543	5.314	1
12	14.03	14.16	0.8796	5.438	3.201	1.717	5.001	1
13	13.89	14.02	0.888	5.439	3.199	3.986	4.738	1
14	13.78	14.06	0.8759	5.479	3.156	3.136	4.872	1
15	13.74	14.05	0.8744	5.482	3.114	2.932	4.825	1
16	14.59	14.28	0.8993	5.351	3.333	4.185	4.781	1
17	13.99	13.83	0.9183	5.119	3.383	5.234	4.781	1
18	15.69	14.75	0.9058	5.527	3.514	1.599	5.046	1
19	14.7	14.21	0.9153	5.205	3.466	1.767	4.649	1
20	12.72	13.57	0.8686	5.226	3.049	4.102	4.914	1
21	14.16	14.4	0.8584	5.658	3.129	3.072	5.176	1
22	14.11	14.26	0.8722	5.52	3.168	2.688	5.219	1
23	15.88	14.9	0.8988	5.618	3.507	0.7651	5.091	1
24	12.08	13.23	0.8664	5.099	2.936	1.415	4.961	1
25	15.01	14.76	0.8657	5.789	3.245	1.791	5.001	1

— Tweak both algorithms  
4. for best results.

— Compare the two  
5. algorithms.

— Visualize the obtained  
clusters in a 2D scatter plot.  
6.

— Reduce the dimensionality  
7. for better visualization.

8. Submit :-

Report

- Code without data set.
- Text file on how to run the code.
- Between 1000 - 5000 characters of report. Include visualization of 2D scatter plot.
- How the clustering algorithms comparatively performed. State opinions on why the results were such.