COURSEWORK 2

Data Minning and Visualisation

Usman Shoukat
201537600

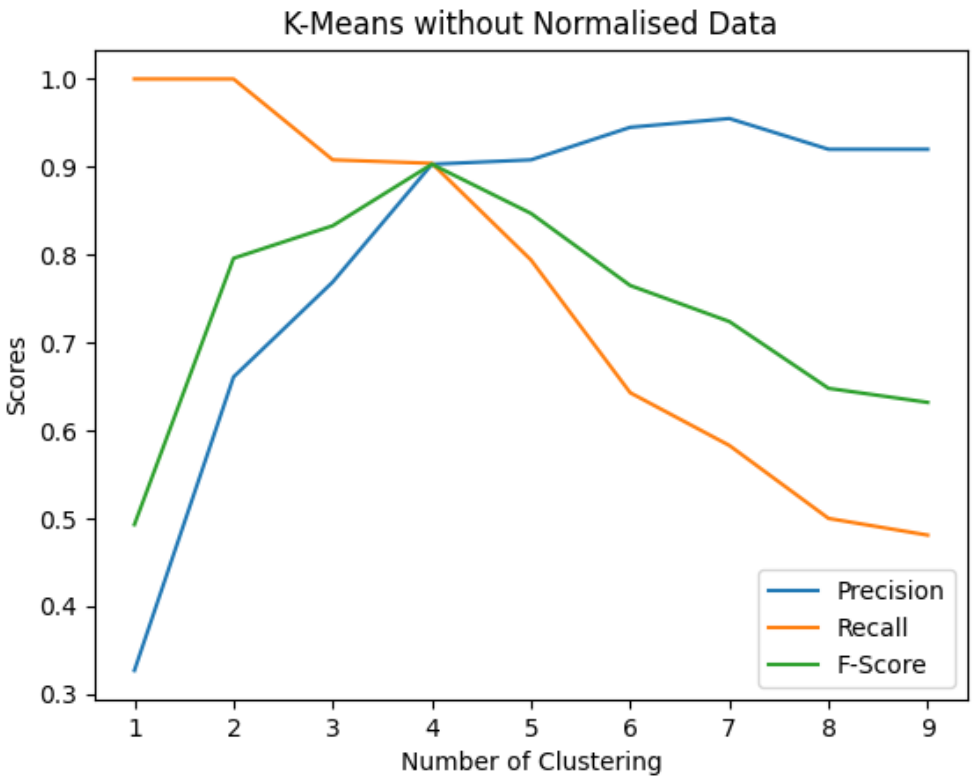
Q3

The results with k-means clustering without normalisation and seed value 45 are as follows:

Clustering Evaluation

---

k	Precision	Recall	F-score
1	0.327	1.0	0.493
2	0.661	1.0	0.796
3	0.769	0.908	0.833
4	0.903	0.904	0.903
5	0.908	0.794	0.847
6	0.945	0.643	0.765
7	0.955	0.583	0.724
8	0.92	0.5	0.648
9	0.92	0.481	0.632



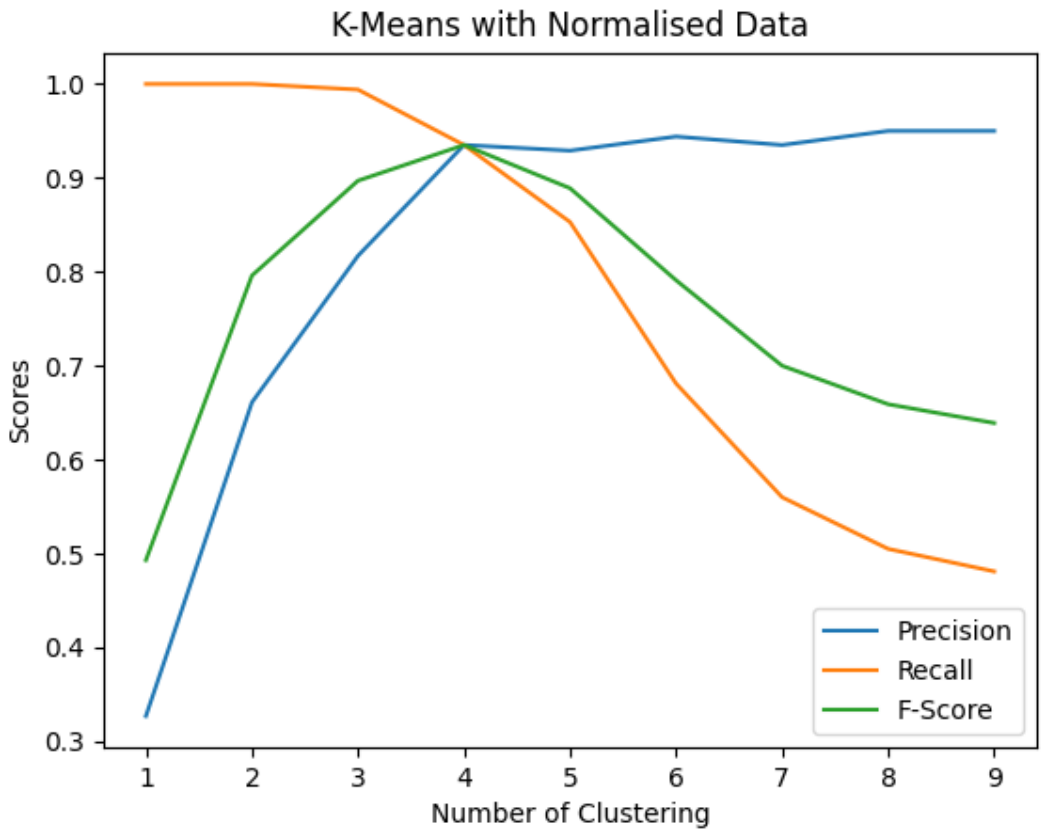
Q4

The results with k-means clustering with normalisation and seed value 45 are as follows:

Clustering Evaluation

---

k	Precision	Recall	F-score
1	0.327	1.0	0.493
2	0.661	1.0	0.796
4	0.935	0.935	0.935
5	0.929	0.853	0.889
6	0.944	0.681	0.791
7	0.935	0.56	0.7
8	0.95	0.505	0.659
9	0.95	0.481	0.639



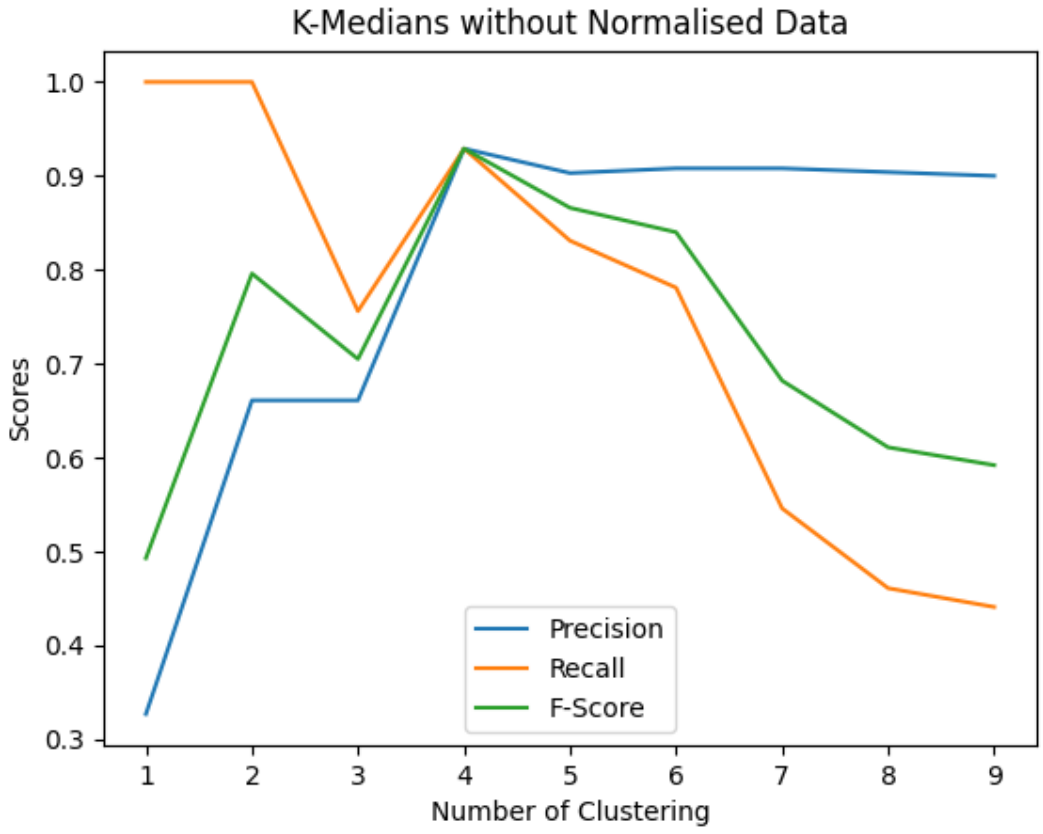
Q5

The results with k-medians clustering without normalisation and seed value 10 are as follows:

Clustering Evaluation

---

k	Precision	Recall	F-score
1	0.327	1.0	0.493
2	0.661	1.0	0.796
3	0.661	0.756	0.705
4	0.929	0.929	0.929
5	0.903	0.831	0.866
6	0.908	0.781	0.84
7	0.908	0.546	0.682
8	0.904	0.461	0.611
9	0.9	0.441	0.592



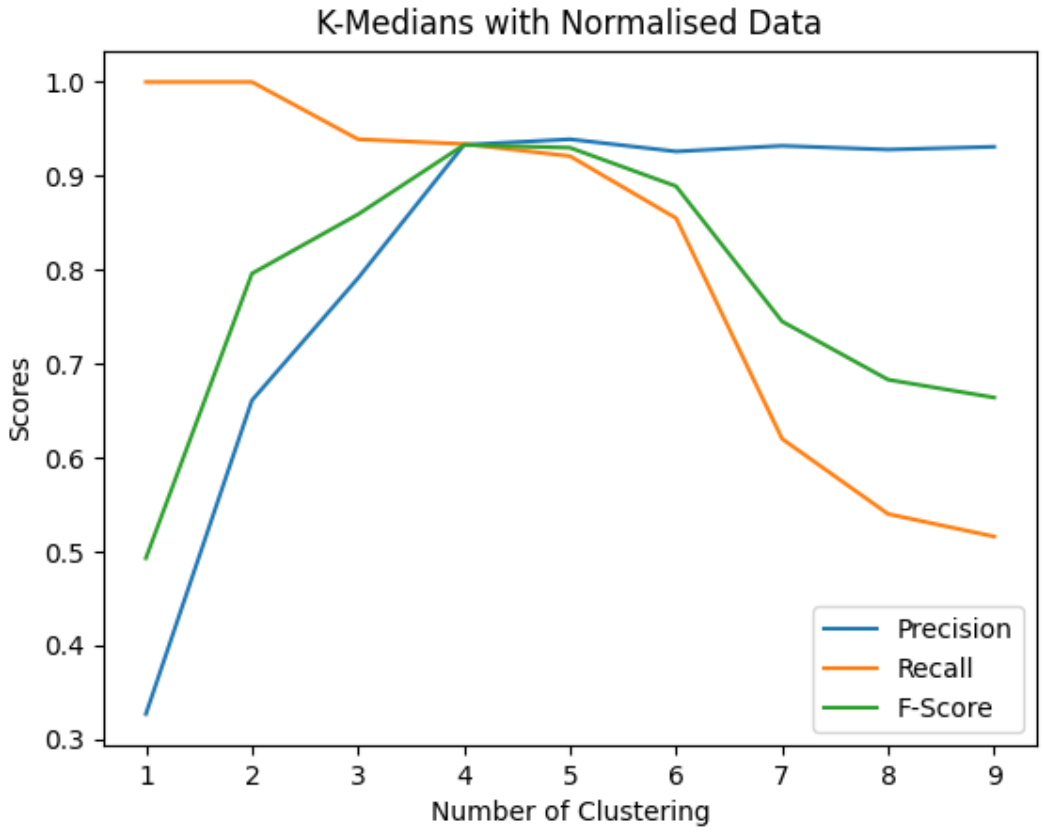
Q6

The results with k-medians clustering with normalisation and seed value 10 are as follows:

Clustering Evaluation

---

k	Precision	Recall	F-score
1	0.327	1.0	0.493
2	0.661	1.0	0.796
3	0.791	0.939	0.859
4	0.933	0.934	0.933
5	0.939	0.921	0.93
6	0.926	0.855	0.889
7	0.932	0.62	0.745
8	0.928	0.54	0.683
9	0.931	0.516	0.664



Q7

From the analysis of all the above four settings, I can see that the precision and recall curves intersect at k= 4. Likewise, I have the highest f-score for all four settings at k = 4. Hence, the best number of clustering is 4.

If I compare the different settings among each other, there is no significant change between the results of different settings. But I can see that k-means with normalisation gives the smoothened curves and highest f-score at k = 4 among all others. This justifies that **k-means with normalisation** is the best setting for the given dataset.