CENG499

Assignment 2

Berke Sina Ahlatcı

2468502

Contents

List of Figures

1 Part 1

Following figure contains a table of hyperparameter values I've tested and their accuracy scores on test indices.

K and Similarity Function	Accuracy and Confidence Interval
K = 5, similarity_function = cosine	94.13333333333333 ± 11.892628897922608
K = 5, similarity_function = minkowski	94.13333333333334 ± 11.892628897922597
K = 5, similarity_function = mahalanobis	90.26666666666668 ± 15.000472628603445
K = 10, similarity_function = cosine	95.6 ± 10.379244108636561
K = 10, similarity_function = minkowski	95.06666666666668 ± 10.959602566462644
K = 10, similarity_function = mahalanobis	86.26666666666665 ± 17.41886812158119
K = 15, similarity_function = cosine	94.4 ± 11.635644729307721
K = 15, similarity_function = minkowski	95.86666666666667 ± 10.073826586938225
K = 15, similarity_function = mahalanobis	84.93333333333334 ± 18.103315337047864

Figure 1: Accuracy values with confidence intervals.

I've calculated the accuracies 5 times for each configuration and picked the best hyperparameter configuration by comparing their accuracy scores. Best configuration :

- K : 15
- Similarity Function : Minkowski Distance
- Accuracy : 95.87% with confidence interval of 10.07

2 Part 2

2.1 Kmeans

Following plots display loss values of different K values for datasets one and two.

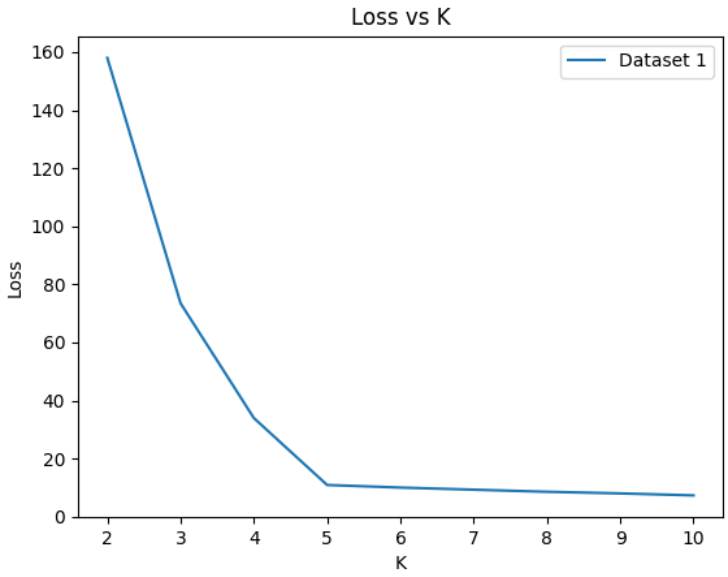

Figure 2: K versus Loss graph of Dataset 1 (Kmeans).

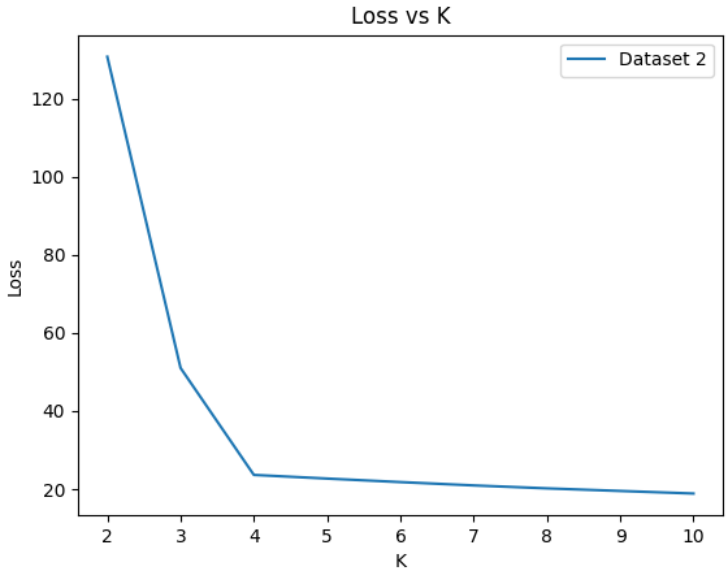

Figure 3: K versus Loss graph of Dataset 2 (Kmeans).

With the help of the elbow method, we can choose following K values for the best outcome:

- For dataset 1, best K value is 5.
- For dataset 2, best K value is 4.

2.2 Kmeans++

K versus loss graphs of kmeans++ algorithm.

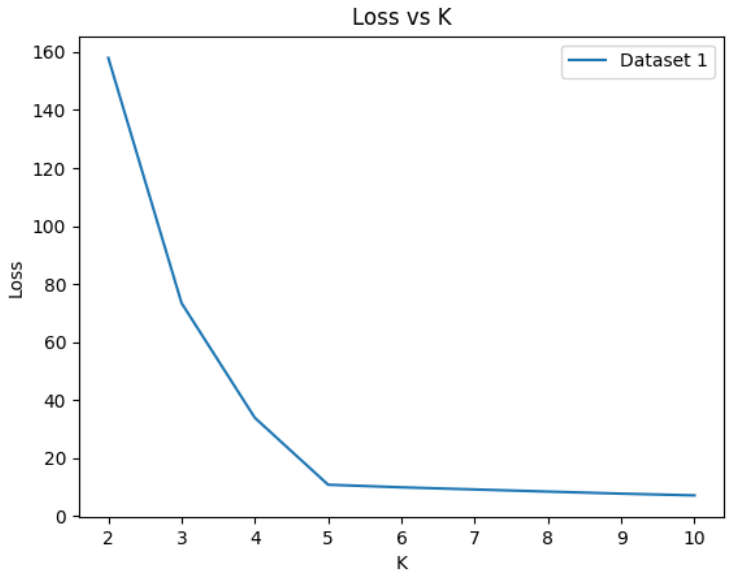

Figure 4: K versus Loss graph of Dataset 1 (Kmeans++).

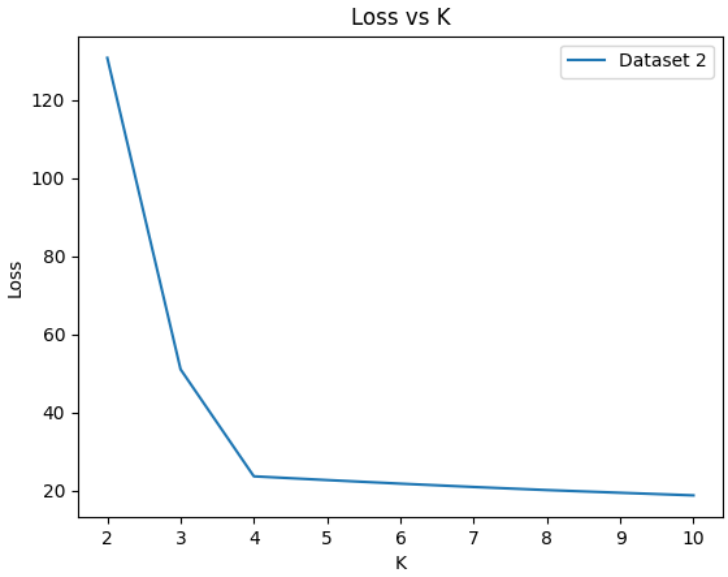

Figure 5: K versus Loss graph of Dataset 2 (Kmeans++).

With the help of the elbow method, we can choose following K values as the most suitable cluster numbers:

- For dataset 1, best K value is 5.
- For dataset 2, best K value is 4.

3 Part 3

Following subsections contain silhouette values and dendrogram plots of each configuration with their optimum K values. I've chosen optimum K values by comparing their silhouette scores.

3.1 Linkage: Single, Similarity: Euclidean

Optimum K value for this configuration is 2.

- K=2

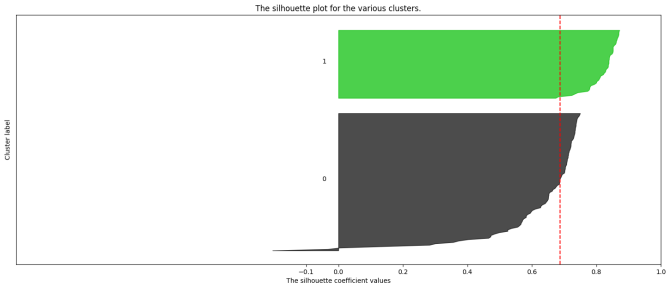

Figure 6: Silhouette plot for K value of 2.

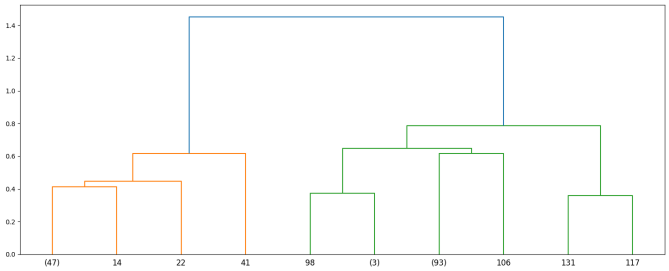

Figure 7: Dendrogram plot for K value of 2.

- K=3

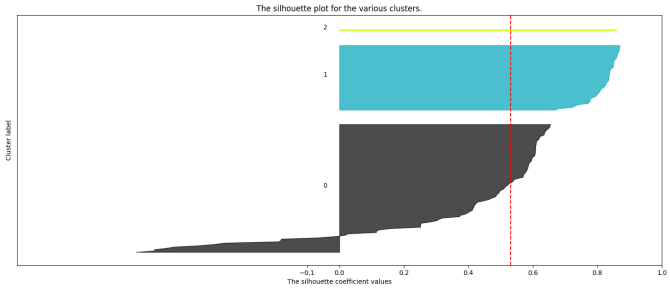

Figure 8: Silhouette plot for K value of 3.

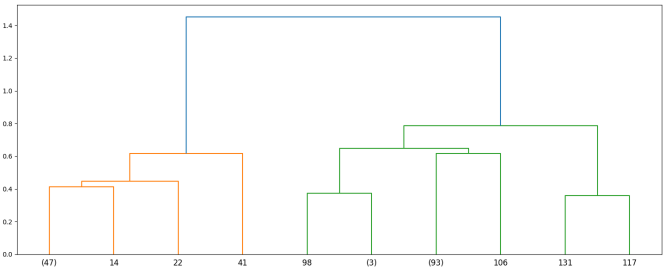

Figure 9: Dendrogram plot for K value of 3.

- K=4

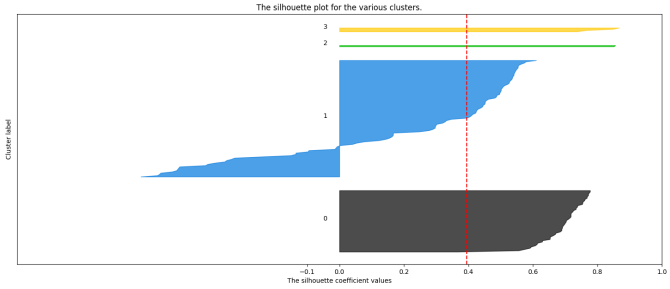

Figure 10: Silhouette plot for K value of 4.

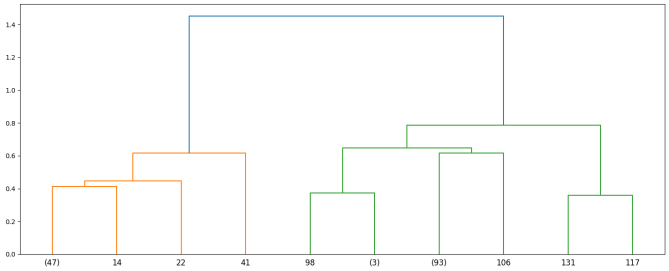

Figure 11: Dendrogram plot for K value of 4.

- K=5

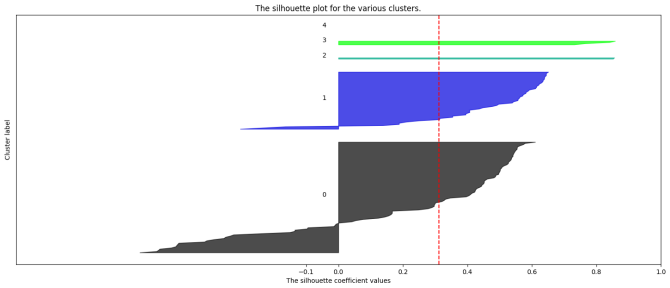

Figure 12: Silhouette plot for K value of 5.

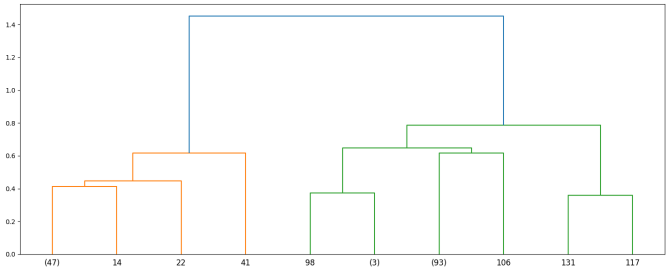

Figure 13: Dendrogram plot for K value of 5.

3.2 Linkage: Single, Similarity: Cosine

Optimum K value for this configuration is 2.

- K=2

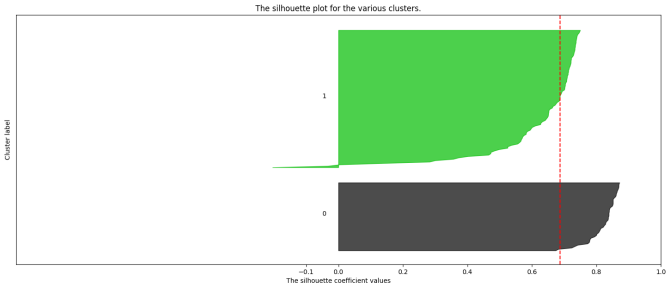

Figure 14: Silhouette plot for K value of 2.

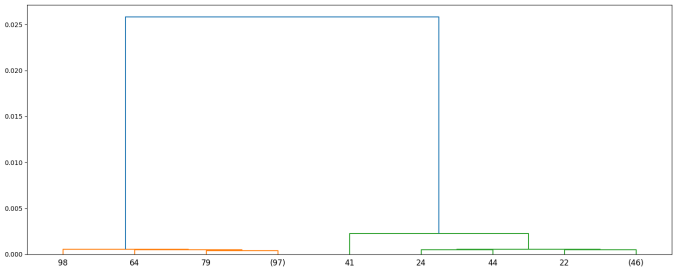

Figure 15: Dendrogram plot for K value of 2.

- K=3

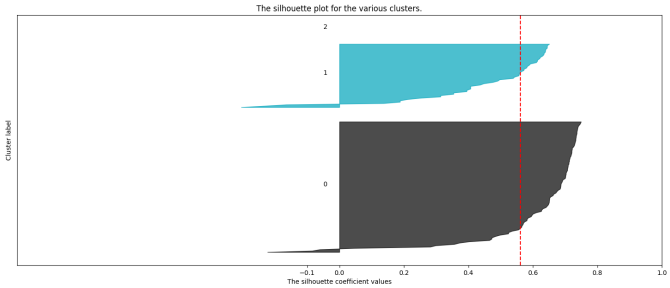

Figure 16: Silhouette plot for K value of 3.

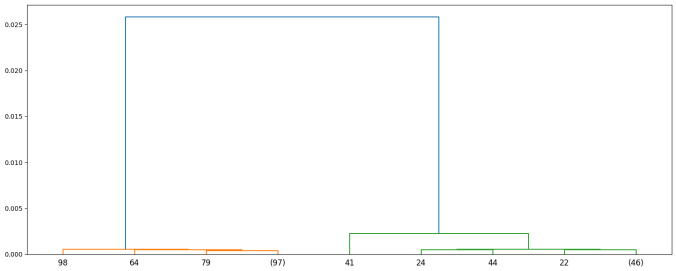

Figure 17: Dendrogram plot for K value of 3.

- K=4

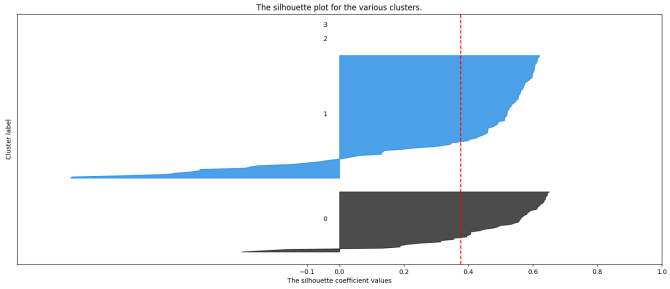

Figure 18: Silhouette plot for K value of 4.

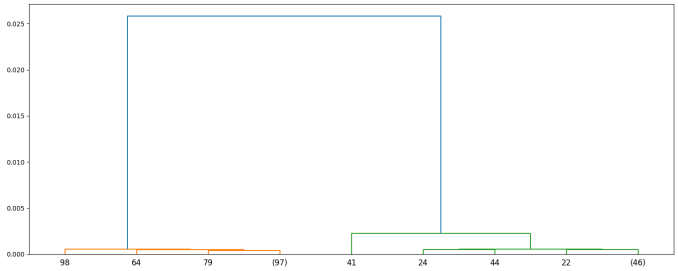

Figure 19: Dendrogram plot for K value of 4.

- K=5

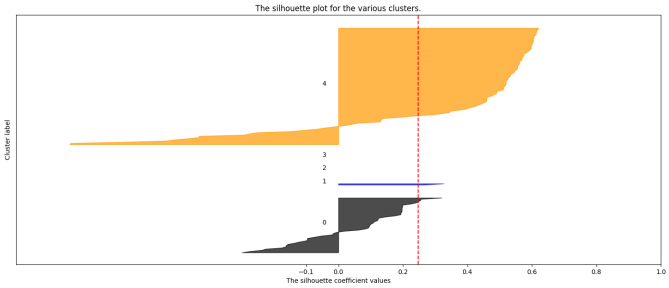

Figure 20: Silhouette plot for K value of 5.

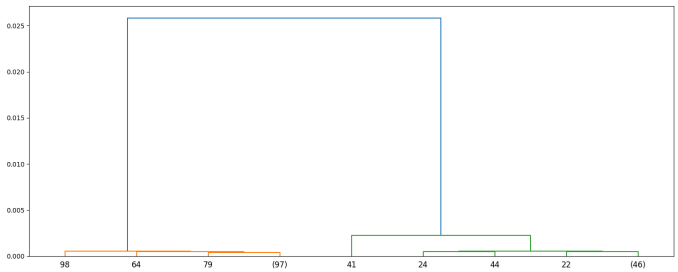

Figure 21: Dendrogram plot for K value of 5.

3.3 Linkage: Complete, Similarity: Euclidean

Optimum K value for this configuration is 4.

- K=2

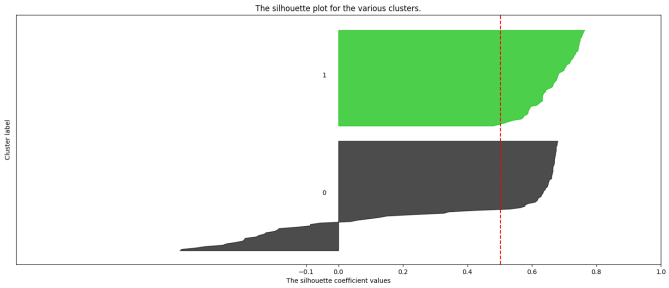

Figure 22: Silhouette plot for K value of 2.

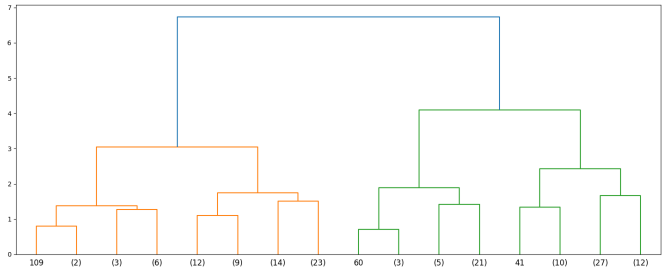

Figure 23: Dendrogram plot for K value of 2.

- K=3

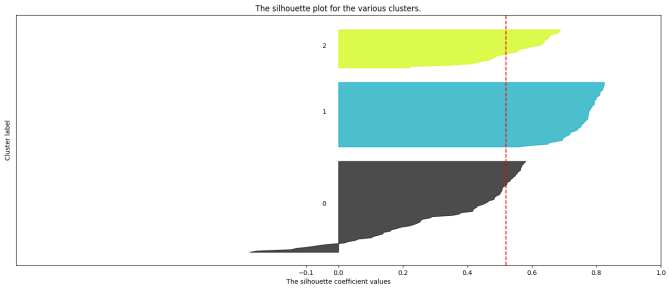

Figure 24: Silhouette plot for K value of 3.

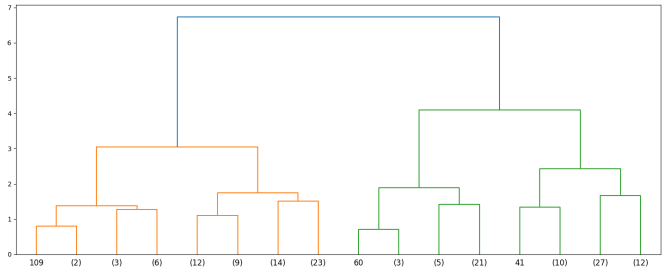

Figure 25: Dendrogram plot for K value of 3.

- K=4

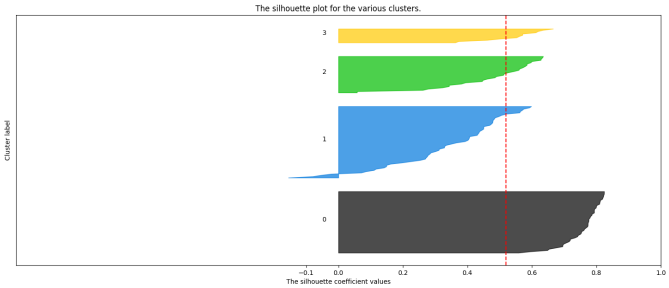

Figure 26: Silhouette plot for K value of 4.

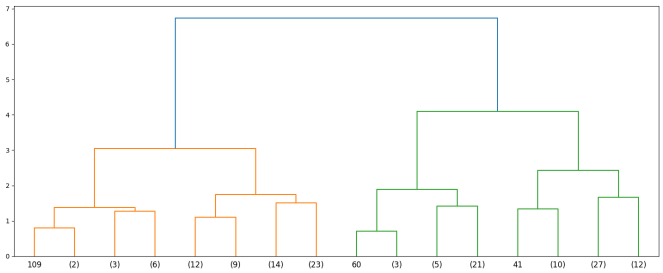

Figure 27: Dendrogram plot for K value of 4.

- K=5

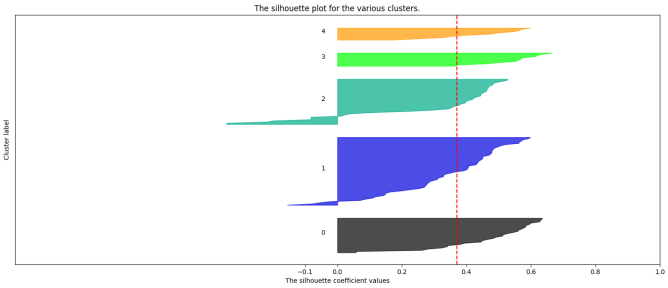

Figure 28: Silhouette plot for K value of 5.

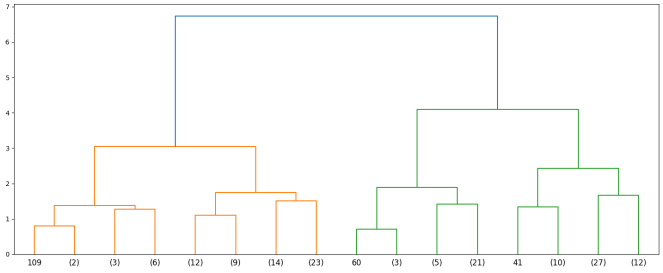

Figure 29: Dendrogram plot for K value of 5.

3.4 Linkage: Complete, Similarity: Cosine

Optimum K value for this configuration is 2.

- K=2

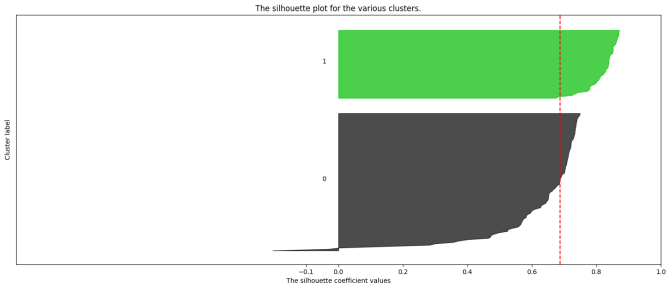

Figure 30: Silhouette plot for K value of 2.

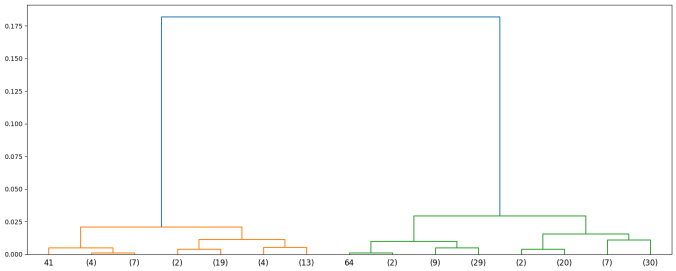

Figure 31: Dendrogram plot for K value of 2.

- K=3

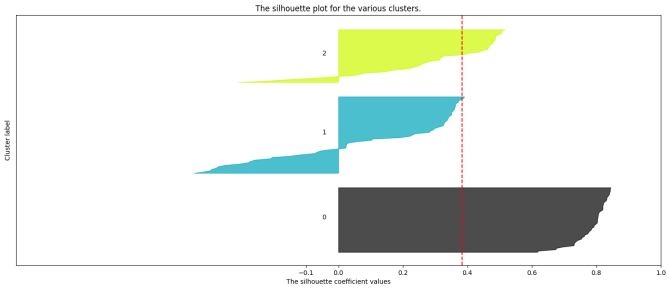

Figure 32: Silhouette plot for K value of 3.

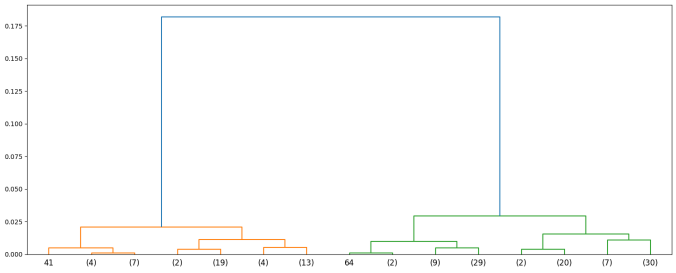

Figure 33: Dendrogram plot for K value of 3.

- K=4

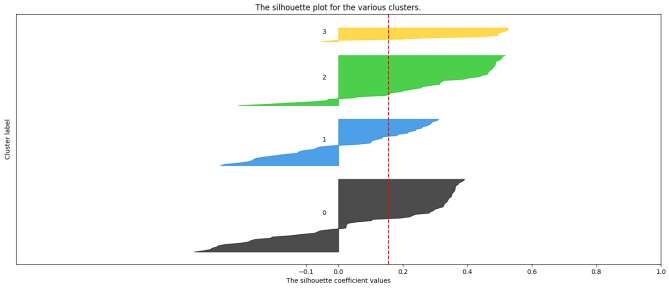

Figure 34: Silhouette plot for K value of 4.

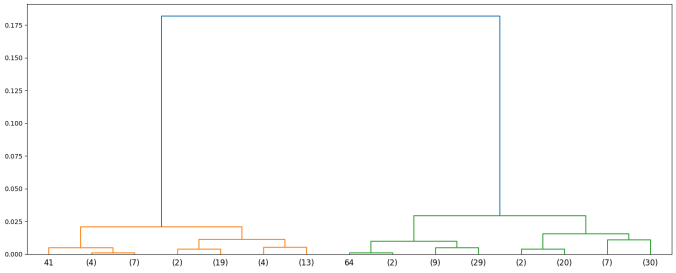

Figure 35: Dendrogram plot for K value of 4.

- K=5

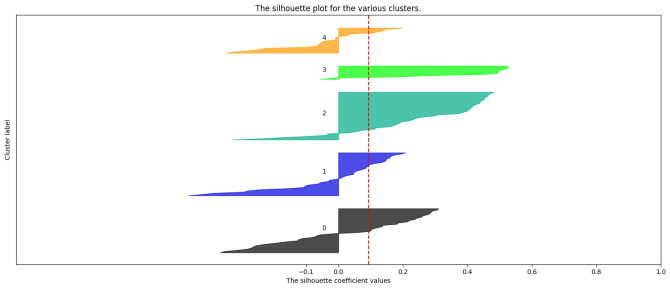

Figure 36: Silhouette plot for K value of 5.

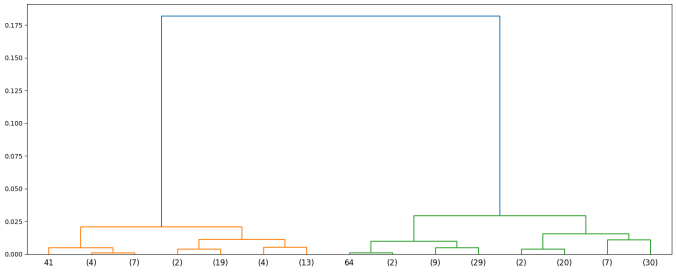

Figure 37: Dendrogram plot for K value of 5.