

BAB III METODOLOGI PENELITIAN

3.1 Jenis Data

Dataset yang digunakan bersumber dari *UCI Machine Learning Repository*. *Dataset* tersebut merupakan *dataset* yang telah ternormalisasi, teruji dan valid sehingga dapat dipercaya sebagai sumber data pada penelitian ini. Adapun informasi dari *dataset* yang digunakan adalah *seed dataset*. *Seed dataset* atau *dataset* biji gandum merupakan data 210 biji gandum yang dipilih secara acak dan memiliki tujuh atribut geometrik yaitu *Area A*, *Perimeter P*, *Compactness C = 4*pi*A/p^2*, *Length of Kernel*, *Width of Kernel*, *Asymmetry Coefficient*, *Length of Kernel Groove*. Yang memiliki karakteristik data Multivariansi dan data dapat dijadikan *sample* data untuk metode *Classification*, *Clustering*.

Tabel 3.1 *Dataset Seeds*

NO	Area A	Perimeter P	Compactness C	Length of Kernel	Width of Kernel	Asymmetry Coefficient	Length of Lernel Groove
1	15.26	14.84	0.871	5.763	3.312	2.221	5.22
2	14.88	14.57	0.8811	5.554	3.333	1.018	4.956
3	14.29	14.09	0.905	5.291	3.337	2.699	4.825
4	13.84	13.94	0.8955	5.324	3.379	2.259	4.805
5	16.14	14.99	0.9034	5.658	3.562	1.355	5.175
6	14.38	14.21	0.8951	5.386	3.312	2.462	4.956
7	14.69	14.49	0.8799	5.563	3.259	3.586	5.219
8	14.11	14.1	0.8911	5.42	3.302	2.7	5
9	16.63	15.46	0.8747	6.053	3.465	2.04	5.877
10	16.44	15.25	0.888	5.884	3.505	1.969	5.533
11	15.26	14.85	0.8696	5.714	3.242	4.543	5.314
12	14.03	14.16	0.8796	5.438	3.201	1.717	5.001
13	13.89	14.02	0.888	5.439	3.199	3.986	4.738
14	13.78	14.06	0.8759	5.479	3.156	3.136	4.872
15	13.74	14.05	0.8744	5.482	3.114	2.932	4.825
16	14.59	14.28	0.8993	5.351	3.333	4.185	4.781
17	13.99	13.83	0.9183	5.119	3.383	5.234	4.781
18	15.69	14.75	0.9058	5.527	3.514	1.599	5.046
19	14.7	14.21	0.9153	5.205	3.466	1.767	4.649
20	12.72	13.57	0.8686	5.226	3.049	4.102	4.914
21	14.16	14.4	0.8584	5.658	3.129	3.072	5.176
22	14.11	14.26	0.8722	5.52	3.168	2.688	5.219
23	15.88	14.9	0.8988	5.618	3.507	0.7651	5.091
24	12.08	13.23	0.8664	5.099	2.936	1.415	4.961
25	15.01	14.76	0.8657	5.789	3.245	1.791	5.001

Dataset biji gandum kemudian dijadikan bahan eksperimen di *Institute of Agrophysics of the Polish Academy* di Lublin, Polandia dan di publish di *UCI Machine Learning Respository* dan *link* data sebagai berikut

<https://archive.ics.uci.edu/ml/datasets/seeds>. Sebagai contoh *dataset Seed* dapat dilihat di Tabel 3.1 yang menampilkan 25 data pertama dan lebih lengkapnya di lembar Lampiran 1.

3.2 Metode Rank Order Centroid (ROC)

3.2.1 Pembobotan Atribut dengan Metode ROC

Penentuan *centroid* awal dalam proses *clustering* K-Means dengan pemberian pembobotan setiap atribut menggunakan metode ROC. Penelitian ini memprioritaskan atribut *Area A* di urutan pertama kemudian atribut *Perimeter P* urutan kedua kemudian atribut *Compactness C* urutan ketiga kemudian *Length of Kernel* urutan keempat kemudian atribut *Width of Kernel* urutan kelima kemudian atribut *Asymmetry Coefficient* urutan keenam kemudian atribut *Length of Kernel Groove* urutan ketujuh. Pembobotan ini dikarenakan penelitian ini bertujuan untuk mengetahui pengelompokan biji berdasarkan area dari biji gandum yang digunakan sebagai acuan pengelompokan petani dalam memilih biji gandum. Pembobotan atribut menggunakan metode ROC dengan persamaan (2,3).

Perhitungan pembobotan setiap atribut :

Area A

$$W1 = \frac{\left(\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7}\right)}{7} = 0.37000$$

Perimeter P

$$W2 = \frac{\left(\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7}\right)}{7} = 0.23000$$

Compactness C

$$W3 = \frac{\left(\frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7}\right)}{7} = 0.16000$$

Length of Kernel

$$W4 = \frac{\left(\frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7}\right)}{7} = 0.11000$$

Width of Kernel

$$W5 = \frac{\left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7}\right)}{7} = 0.07000$$

Asymmetry Coefficient

$$W6 = \frac{\left(\frac{1}{6} + \frac{1}{7}\right)}{7} = 0.04000$$

Length of Kernel Groove

$$W7 = \frac{\left(\frac{1}{7}\right)}{7} = 0.02000$$

3.2.2 Penentuan *centroid* Awal dengan Metode ROC

Setelah nilai bobot setiap atribut didapatkan, selanjutnya melakukan perkalian nilai bobot setiap atribut dengan atribut data. Hasil akhir dalam perhitungan ini untuk menentukan data yang akan digunakan sebagai *centroid* awal di metode K-Means. Sebagai contoh perhitungan data pertama menggunakan persamaan (2,2) dapat dilihat sebagai berikut :

Data ke-1

$$\text{Area A} : 15.26000 \times 0.37041 = 5.65243$$

$$\text{Perimeter P} : 14.84000 \times 0.22755 = 3.37686$$

$$\text{Compactness C} : 0.87100 \times 0.15612 = 0.13598$$

$$\text{Length of kernel} : 5.76300 \times 0.10850 = 0.62531$$

$$\text{Width of kernel} : 3.31200 \times 0.07279 = 0.24108$$

$$\text{Asymmetry coefficient} : 2.22100 \times 0.04422 = 1.22679$$

$$\text{Length of kernel groove} : 5.22000 \times 0.02041 = 0.23082$$

$$\begin{aligned} \text{ROC} &= \left(\frac{1}{5.65243} + \frac{1}{3.37686} + \frac{1}{0.13598} + \frac{1}{0.62531} + \frac{1}{0.24108} + \frac{1}{1.22679} + \frac{1}{0.23082} \right) \\ &= 18.72177312 \end{aligned}$$

Penelitian ini melakukan metode *clustering* menggunakan 2 *cluster*, 3 *cluster*, 4 *cluster* dan 5 *cluster*, sehingga nilai akhir ROC yang dibutuhkan adalah nilai tertinggi, nilai tertinggi kedua, nilai tengah, nilai terendah, dan nilai tertinggi kedua.

Hasil akhir dari perhitungan ROC yang dipilih menjadi *centroid* awal sebagai berikut nilai tertinggi terdapat di data dengan id 61 dengan nilai 3.030968, nilai terendah terdapat di data dengan id 90 dengan nilai 2.366162, nilai tengah terdapat di data dengan id 3 dengan nilai 2.704775,

nilai tertinggi kedua terdapat di data dengan id 178 dengan nilai 2.979541, sedangkan nilai terendah kedua terdapat di data dengan id 89 dengan nilai 2.37371.

3.3 Metode *Clustering* K-Means dan ROC

3.3.1 Dua *Cluster*

Nilai *centroid* dari perhitungan ROC yang diambil sebagai *centroid* awal untuk dua *cluster* adalah nilai yang memiliki nilai tertinggi dan terendah. Nilai yang terpilih untuk *centroid* awal adalah data dengan id 61 dan dengan id 90 atau bisa dilihat pada Tabel 3.2.

Tabel 3.2 *Centroid* awal dua *cluster* dengan ROC.

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	11.42000	12.86000	0.86830	5.00800	2.85000	2.70000	4.00000
2	20.88000	17.05000	0.90310	6.45000	4.03200	5.01600	6.32100

Langkah selanjutnya mencari *euclidean distance* dari setiap data *Seeds* dengan *centroid* awal yang telah ditentukan menggunakan persamaan (2.2) dan menggunakan *centroid* yang di Tabel 3.2. Hasil perhitungan data pertama di iterasi pertama menggunakan *Euclidean distance* dapat dilihat di Tabel 3.3.

Tabel 3.3 Hasil perhitungan di iterasi pertama untuk dua *cluster*

Cluster	Hasil
1	4.600800
2	6.81793

Hasil pengelompokkan menggunakan hasil perhitungan *euclidean distance* di iterasi pertama dapat dilihat di tabel 3.4 yang menampilkan contoh hasil akhir 25 data pertama. Penentuan suatu data masuk ke *cluster* pertama atau kedua adalah dengan mencari hasil akhir perhitungan *euclidean distance* yang terkecil antar *cluster*. Sehingga dari iterasi pertama

didapatkan jumlah anggota *cluster* pertama 141 data sedangkan *cluster* kedua 69 data.

Tabel 3.4 Hasil mengelompokkan dua *cluster* di iterasi pertama

NO	C1	C2	TERDEKAT	CLUSTER
1	4.60080	6.81793	4.60080	C1
2	4.37839	7.82876	4.37839	C1
3	3.27857	7.85000	3.27857	C1
4	2.87154	8.41590	2.87154	C1
5	5.56198	6.50197	5.56198	C1
6	3.45134	7.76867	3.45134	C1
7	4.01201	7.03661	4.01201	C1
8	3.18563	7.95225	3.18563	C1
9	6.27161	5.48855	5.48855	C2
10	5.91555	5.78421	5.78421	C2
11	4.94783	6.23130	4.94783	C1
12	3.28308	8.34360	3.28308	C1
13	3.15482	7.95764	3.15482	C1
14	2.87674	8.16670	2.87674	C1
15	2.79774	8.26688	2.79774	C1
16	3.90287	7.21080	3.90287	C1
17	3.85692	7.90270	3.85692	C1
18	4.98227	6.82992	4.98227	C1
19	3.78062	7.84093	3.78062	C1
20	2.25438	9.16388	2.25438	C1
21	3.44978	7.66262	3.44978	C1
22	3.32348	7.86533	3.32348	C1
23	5.45820	7.08332	5.45820	C1
24	1.77846	10.48222	1.77846	C1
25	4.36948	7.27336	4.36948	C1

Langkah selanjutnya mencari *centroid* baru untuk iterasi kedua dengan membagi jumlah semua data tiap *cluster* dengan jumlah anggota *cluster* di tiap atribut *dataset* atau menggunakan persamaan (2,1). Contoh perhitungan di *cluster* pertama sebagai berikut :

$$V_{Area A} = \frac{1845.93000}{141} = 13.09170$$

$$V_{Perimeter P} = \frac{1941.28000}{141} = 13.76794$$

$$V_{Compactness C} = \frac{121.89620}{141} = 0.86451$$

$$V_{\text{Length of Kernel}} = \frac{756.61300}{141} = 5.36605$$

$$V_{\text{Width of Kernel}} = \frac{429.57700}{141} = 3.04665$$

$$V_{\text{Asymmetry Coefficient}} = \frac{527.99020}{141} = 3.74461$$

$$V_{\text{Length of Kernel Groove}} = \frac{718.51000}{141} = 5.09582$$

Hasil perhitungan *centroid cluster* kedua juga menggunakan persamaan (2,1) dan hasil akhir perhitungan pencarian *centroid* baru dapat dilihat di Tabel 3.5.

Tabel 3.5 Hasil *centroid* baru dua *cluster* di iterasi kedua

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	13.09170	13.76794	0.86451	5.36605	3.04665	3.74461	5.09582
2	18.43551	16.17638	0.88425	6.16491	3.69174	3.60554	6.01657

Langkah selanjutnya mengelompokkan kembali dengan persamaan (2,2) menggunakan *centroid* baru. Hasil perhitungan data pertama di iterasi kedua menggunakan *centroid* baru dapat dilihat di Tabel 3.6.

Tabel 3.6 Hasil perhitungan dua *cluster* di iterasi kedua

Cluster	Hasil
1	2.90098
2	3.83759

Hasil akhir perhitungan *euclidean distance* di iterasi kedua dapat dilihat di tabel 3.7 yang menampilkan contoh hasil akhir 25 data pertama. Penentuan suatu data masuk ke *cluster* pertama atau kedua adalah dengan mencari hasil akhir perhitungan *euclidean distance* yang terdekat antar *cluster*. Sehingga dari iterasi kedua didapatkan jumlah anggota untuk *cluster* pertama 137 data sedangkan *cluster* kedua 73 data. Dikarenakan anggota *cluster* sebelumnya dengan anggota *cluster* baru tidak sama atau

berubah, maka proses *clustering* akan dilanjutkan dengan mengulangi kembali proses *clustering* dalam pencarian *centroid* baru menggunakan anggota *cluster* terakhir.

Tabel 3.7 Hasil mengelompokkan dua *cluster* di iterasi kedua

NO	C1	C2	TERDEKAT	CLUSTER
1	2.90098	3.83759	2.90098	C1
2	3.37829	4.85223	3.37829	C1
3	1.67268	4.96687	1.67268	C1
4	1.73042	5.49602	1.73042	C1
5	4.10541	3.56709	3.56709	C2
6	1.89513	4.84740	1.89513	C1
7	1.78896	4.24956	1.78896	C1
8	1.52198	5.05782	1.52198	C1
9	4.42104	2.51147	2.51147	C2
10	4.15096	2.80466	2.80466	C2
11	2.59165	3.69071	2.59165	C1
12	2.27674	5.37042	2.27674	C1
13	0.95718	5.27834	0.95718	C1
14	1.00220	5.33300	1.00220	C1
15	1.11875	5.40768	1.11875	C1
16	1.69811	4.58671	1.69811	C1
17	1.81803	5.53507	1.81803	C1
18	3.54510	3.87034	3.54510	C1
19	2.66402	4.90351	2.66402	C1
20	0.59813	6.49748	0.59813	C1
21	1.44625	4.79593	1.44625	C1
22	1.56497	4.95488	1.56497	C1
23	4.26735	4.17348	4.17348	C2
24	2.61567	7.52954	2.61567	C1
25	2.95091	4.29015	2.95091	C1

Langkah mencari *centroid* baru untuk iterasi ketiga dengan menggunakan persamaan (2,1). Contoh perhitungan *cluster* pertama sebagai berikut :

$$V_{Area A} = \frac{1781.39000}{137} = 13.00285$$

$$V_{\text{Perimeter P}} = \frac{1881.25000}{137} = 13.73175$$

$$V_{\text{Compactness C}} = \frac{118.29500}{137} = 0.86347$$

$$V_{\text{Length of Kernel}} = \frac{734.10600}{137} = 5.35844$$

$$V_{\text{Width of Kernel}} = \frac{415.45400}{137} = 3.03251$$

$$V_{\text{Asymmetry Coefficient}} = \frac{525.22410}{137} = 3.83375$$

$$V_{\text{Length of Kernel Groove}} = \frac{697.48600}{137} = 5.09114$$

Perhitungan *centroid cluster* kedua juga menggunakan persamaan (2,1) dan hasil akhir perhitungan pencarian *centroid* baru dapat dilihat di Tabel 3.8.

Tabel 3.8 Hasil *centroid* baru dua *cluster* di iterasi ketiga

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	13.00285	13.73175	0.86347	5.35844	3.03251	3.83375	5.09114
2	18.30945	16.11233	0.88513	6.13542	3.68292	3.44586	5.97489

Langkah selanjutnya mengelompokkan kembali dengan persamaan (2,2) menggunakan *centroid* baru. Hasil perhitungan data pertama di iterasi ketiga menggunakan *centroid* baru dapat dilihat di Tabel 3.9.

Tabel 3.9 Hasil perhitungan di iterasi ketiga untuk dua *cluster*

Cluster	Hasil
1	3.03024
2	3.64206

Hasil akhir perhitungan *euclidean distance* di iterasi ketiga dapat dilihat di Tabel 3.10 yang menampilkan contoh hasil akhir 25 data pertama. Penentuan suatu data masuk ke *cluster* pertama atau kedua adalah dengan

mencari hasil akhir perhitungan *euclidean distance* yang terdekat antar *cluster*. Sehingga dari iterasi ketiga didapatkan jumlah anggota *cluster* pertama 134 data sedangkan *cluster* kedua 76 data.

Tabel 3.10 Hasil mengelompokkan dua *cluster* di iterasi ketiga

NO	C1	C2	TERDEKAT	CLUSTER
1	3.03024	3.64206	3.03024	C1
2	3.50741	4.64038	3.50741	C1
3	1.80072	4.79155	1.80072	C1
4	1.85153	5.31255	1.85153	C1
5	4.23649	3.34971	3.34971	C2
6	2.02611	4.66529	2.02611	C1
7	1.89541	4.10254	1.89541	C1
8	1.65296	4.88350	1.65296	C1
9	4.54398	2.29924	2.29924	C2
10	4.27688	2.59032	2.59032	C2
11	2.65862	3.59240	2.65862	C1
12	2.40046	5.17508	2.40046	C1
13	1.02607	5.14298	1.02607	C1
14	1.12984	5.17194	1.12984	C1
15	1.24527	5.24120	1.24527	C1
16	1.76941	4.46110	1.76941	C1
17	1.79560	5.44184	1.79560	C1
18	3.67639	3.65928	3.65928	C2
19	2.79108	4.70502	2.79108	C1
20	0.47680	6.36322	0.47680	C1
21	1.57229	4.63287	1.57229	C1
22	1.69660	4.78083	1.69660	C1
23	4.39923	3.95457	3.95457	C2
24	2.65468	7.34157	2.65468	C1
25	3.08180	4.08832	3.08180	C1

Dikarenakan anggota *cluster* sebelumnya dengan anggota *cluster* baru tidak sama atau berubah, maka proses *clustering* akan dilanjutkan dengan mengulangi kembali proses *clustering* dalam pencarian *centroid* baru menggunakan anggota *cluster* terakhir. Langkah mencari *centroid* baru dengan menggunakan persamaan (2,1). Perhitungan *cluster* pertama dapat dilihat sebagai berikut :

$$V_{\text{Area A}} = \frac{1734.53000}{134} = 12.94425$$

$$V_{\text{Perimeter P}} = \frac{1836.24000}{134} = 13.70328$$

$$V_{\text{Compactness C}} = \frac{115.67850}{134} = 0.86327$$

$$V_{\text{Length of Kernel}} = \frac{716.82700}{134} = 5.34946$$

$$V_{\text{Width of Kernel}} = \frac{405.42300}{134} = 3.02554$$

$$V_{\text{Asymmetry Coefficient}} = \frac{518.26010}{134} = 3.86761$$

$$V_{\text{Length of Kernel Groove}} = \frac{680.80900}{134} = 5.08066$$

Perhitungan *centroid cluster* kedua juga menggunakan persamaan (2,1) dan hasil akhir perhitungan pencarian *centroid* baru dapat dilihat di tabel 3.11.

Tabel 3.11 Hasil *centroid* baru dua *cluster* di iterasi keempat

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	12.94425	13.70328	0.86327	5.34946	3.02554	3.86761	5.08066
2	18.20329	16.06855	0.88462	6.12059	3.66953	3.40148	5.95847

Tabel 3.12 Hasil perhitungan di iterasi keempat untuk dua *cluster*

Cluster	Hasil
1	3.10462
2	3.51668

Langkah selanjutnya mengelompokkan kembali dengan persamaan (2,2) menggunakan *centroid* baru. Hasil perhitungan data pertama di iterasi keempat menggunakan *centroid* baru dapat dilihat di Tabel 3.12. Hasil akhir

perhitungan *euclidean distance* di iterasi keempat dapat dilihat di tabel 3.13 yang menampilkan contoh hasil akhir 25 data pertama.

Tabel 3.13 Hasil mengelompokkan dua *cluster* di iterasi keempat.

NO	C1	C2	TERDEKAT	CLUSTER
1	3.10462	3.51668	3.10462	C1
2	3.57363	4.51775	3.57363	C1
3	1.86916	4.66968	1.86916	C1
4	1.91014	5.18882	1.91014	C1
5	4.30989	3.23237	3.23237	C2
6	2.09614	4.54168	2.09614	C1
7	1.96606	3.98690	1.96606	C1
8	1.72289	4.76098	1.72289	C1
9	4.61880	2.17998	2.17998	C2
10	4.35180	2.46886	2.46886	C2
11	2.71441	3.49493	2.71441	C1
12	2.46126	5.04961	2.46126	C1
13	1.07930	5.03172	1.07930	C1
14	1.19936	5.05234	1.19936	C1
15	1.31136	5.12027	1.31136	C1
16	1.82413	4.35441	1.82413	C1
17	1.80288	5.34733	1.80288	C1
18	3.74893	3.53747	3.53747	C2
19	2.85585	4.58194	2.85585	C1
20	0.40815	6.25112	0.40815	C1
21	1.64665	4.51220	1.64665	C1
22	1.76889	4.65770	1.76889	C1
23	4.47008	3.83997	3.83997	C2
24	2.65919	7.21631	2.65919	C1
25	3.15341	3.96363	3.15341	C1

Penentuan suatu data masuk ke *cluster* pertama atau kedua adalah dengan mencari hasil akhir perhitungan *euclidean distance* yang terdekat antar *cluster*. Sehingga dari iterasi keempat didapatkan jumlah anggota *cluster* pertama 134 data sedangkan *cluster* kedua 76 data. Proses *clustering* berhenti di iterasi keempat karena anggota *cluster* sebelumnya dengan anggota *cluster* baru sama atau tidak mengalami perubahan. Sehingga proses *clustering* 2 *cluster* dengan penentuan *centroid* awal menggunakan

penentuan metode ROC menghasilkan *cluster* pertama dengan anggota 134 dan *cluster* kedua dengan anggota 76.

3.3.2 Tiga Cluster

Nilai *centroid* dari perhitungan ROC yang diambil sebagai *centroid* awal untuk tiga *cluster* adalah nilai yang memiliki nilai tertinggi, nilai terendah, dan nilai tengah. Nilai yang terpilih untuk *centroid* awal adalah data dengan nilai tertinggi di id 61, untuk data dengan nilai terendah di id 90, dan untuk data dengan nilai tengah di id 3 atau bisa dilihat pada Tabel 3.14.

Tabel 3.14 *Centroid* awal tiga *cluster* dengan ROC.

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	11.42000	12.86000	0.86830	5.00800	2.85000	2.70000	4.00000
2	20.88000	17.05000	0.90310	6.45000	4.03200	5.01600	6.32100
3	14.29000	14.09000	0.90500	5.29100	3.33700	2.69900	4.82500

Langkah selanjutnya mencari *euclidean distance* dari setiap data *Seeds* dengan *centroid* awal yang telah ditentukan menggunakan persamaan (2.2) dan menggunakan *centroid* yang di Tabel 3.14. Hasil perhitungan data pertama di iterasi pertama menggunakan *Euclidean distance* dapat dilihat di Tabel 3.15.

Tabel 3.15 Hasil perhitungan di iterasi pertama untuk tiga *cluster*.

Cluster	Hasil
1	4.60080
2	6.81793
3	1.45344

Hasil pengelompokkan menggunakan hasil perhitungan *euclidean distance* di iterasi pertama dapat dilihat di tabel 3.16 yang menampilkan contoh hasil akhir 25 data pertama. Penentuan suatu data masuk ke *cluster* pertama atau kedua atau ketiga adalah dengan mencari hasil akhir perhitungan *euclidean distance* yang terkecil antar *cluster*. Sehingga dari

iterasi pertama didapatkan jumlah anggota *cluster* pertama 64 data , *cluster* kedua 51 data, dan *cluster* ketiga 95 data.

Tabel 3.16 Hasil mengelompokkan tiga *cluster* di iterasi pertama

NO	C1	C2	C3	TERDEKAT	CLUSTER
1	4.60080	6.81793	1.45344	1.45344	C3
2	4.37839	7.82876	1.86847	1.86847	C3
3	3.27857	7.85000	0.00000	0.00000	C3
4	2.87154	8.41590	0.64957	0.64957	C3
5	5.56198	6.50197	2.51926	2.51926	C3
6	3.45134	7.76867	0.32493	0.32493	C3
7	4.01201	7.03661	1.15875	1.15875	C3
8	3.18563	7.95225	0.28493	0.28493	C3
9	6.27161	5.48855	3.08082	3.08082	C3
10	5.91555	5.78421	2.71706	2.71706	C3
11	4.94783	6.23130	2.31239	2.31239	C3
12	3.28308	8.34360	1.05288	1.05288	C3
13	3.15482	7.95764	1.36751	1.36751	C3
14	2.87674	8.16670	0.72328	0.72328	C3
15	2.79774	8.26688	0.66748	0.66748	C3
16	3.90287	7.21080	1.52967	1.52967	C3
17	3.85692	7.90270	2.57248	2.57248	C3
18	4.98227	6.82992	1.93429	1.93429	C3
19	3.78062	7.84093	1.05178	1.05178	C3
20	2.25438	9.16388	2.19092	2.19092	C3
21	3.44978	7.66262	0.74529	0.74529	C3
22	3.32348	7.86533	0.54657	0.54657	C3
23	5.45820	7.08332	2.67035	2.67035	C3
24	1.77846	10.48222	2.73679	1.77846	C1
25	4.36948	7.27336	1.44248	1.44248	C3

Langkah selanjutnya mencari *centroid* baru untuk iterasi kedua dengan membagi jumlah semua data tiap *cluster* dengan jumlah anggota *cluster* di tiap atribut *dataset* atau menggunakan persamaan (2,1). Contoh perhitungan di *cluster* pertama sebagai berikut :

$$V_{Area A} = \frac{747.73000}{64} = 11.68328$$

$$V_{Perimeter P} = \frac{841.66000}{64} = 13.15094$$

$$V_{\text{Compactness C}} = \frac{54.305900}{64} = 0.84853$$

$$V_{\text{Length of Kernel}} = \frac{332.22600}{64} = 5.19103$$

$$V_{\text{Width of Kernel}} = \frac{181.09500}{64} = 2.82961$$

$$V_{\text{Asymmetry Coefficient}} = \frac{288.78600}{64} = 4.51228$$

$$V_{\text{Length of Kernel Groove}} = \frac{322.42500}{64} = 5.03789$$

Hasil perhitungan *centroid cluster* kedua dan ketiga juga menggunakan persamaan (2,1) dan hasil akhir perhitungan pencarian *centroid* baru dapat dilihat di Tabel 3.17.

Tabel 3.17 Hasil *centroid* baru tiga *cluster* di iterasi kedua

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	11.68328	13.15094	0.84853	5.19103	2.82961	4.51228	5.03789
2	19.04569	16.42922	0.88635	6.25506	3.76127	3.56392	6.11200
3	14.72547	14.50421	0.87789	5.58693	3.27776	3.22343	5.25806

Langkah selanjutnya mengelompokkan kembali dengan persamaan (2,2) menggunakan *centroid* baru. Hasil perhitungan data pertama di iterasi kedua menggunakan *centroid* baru dapat dilihat di Tabel 3.18.

Tabel 3.18 Hasil perhitungan di iterasi kedua untuk tiga *cluster*

Cluster	Hasil
1	4.63566
2	4.46098
3	1.19876

Hasil akhir perhitungan *euclidean distance* di iterasi kedua dapat dilihat di tabel 3.19 yang menampilkan contoh hasil akhir 25 data pertama.

Tabel 3.19 Hasil mengelompokkan tiga *cluster* di iterasi kedua

NO	C1	C2	C3	TERDEKAT	CLUSTER
1	4.63566	4.46098	1.19876	1.19876	C3
2	4.98354	5.41319	2.23328	2.23328	C3
3	3.35868	5.62163	0.95687	0.95687	C3
4	3.27525	6.14162	1.52233	1.52233	C3
5	5.83002	4.08266	2.41274	2.41274	C3
6	3.58847	5.49557	0.95853	0.95853	C3
7	3.47102	4.92562	0.36793	0.36793	C3
8	3.21759	5.71642	0.95467	0.95467	C3
9	6.14536	3.04651	2.56450	2.56450	C3
10	5.88929	3.35597	2.29891	2.29891	C3
11	4.02504	4.35700	1.47213	1.47213	C3
12	3.81315	5.99445	1.72211	1.72211	C3
13	2.48823	5.95261	1.34634	1.34634	C3
14	2.70823	6.00251	1.12890	1.12890	C3
15	2.78338	6.07317	1.21975	1.21975	C3
16	3.19037	5.25647	1.13137	1.13137	C3
17	2.58558	6.18962	2.34459	2.34459	C3
18	5.26143	4.43500	1.93253	1.93253	C3
19	4.28015	5.51088	1.66195	1.66195	C3
20	1.21813	7.17501	2.44287	1.21813	C1
21	3.17735	5.46468	0.62277	0.62277	C3
22	3.27156	5.61459	0.86210	0.86210	C3
23	5.94639	4.65903	2.75950	2.75950	C3
24	3.12775	8.16801	3.51224	3.12775	C1
25	4.64686	4.89232	1.51868	1.51868	C3

Penentuan suatu data masuk ke *cluster* pertama atau kedua atau ketiga adalah dengan mencari hasil akhir perhitungan *euclidean distance* yang terdekat antar *cluster*. Sehingga dari iterasi kedua didapatkan jumlah anggota *cluster* pertama 79 data, *cluster* kedua 60 data, dan *cluster* ketiga 71 data. Dikarenakan anggota *cluster* sebelumnya dengan anggota *cluster* baru tidak sama atau berubah, maka proses *clustering* akan dilanjutkan dengan mengulangi kembali proses *clustering* dalam pencarian *centroid* baru menggunakan anggota *cluster* terakhir.

Langkah mencari *centroid* baru dengan menggunakan persamaan (2,1). Perhitungan *cluster* pertama dapat dilihat sebagai berikut :

$$V_{Area A} = \frac{942.06000}{79} = 11.92481$$

$$V_{Perimeter P} = \frac{1047.65000}{79} = 13.26139$$

$$V_{Compactness C} = \frac{67.25530}{79} = 0.85133$$

$$V_{Length\ of\ Kernel} = \frac{412.73800}{79} = 5.22453$$

$$V_{Width\ of\ Kernel} = \frac{226.53000}{79} = 2.86747$$

$$V_{Asymmetry\ Coefficient} = \frac{361.45900}{79} = 4.57543$$

$$V_{Length\ of\ Kernel\ Groove} = \frac{399.83400}{79} = 5.06119$$

Perhitungan *centroid cluster* kedua dan ketiga juga menggunakan persamaan (2,1) dan hasil akhir perhitungan pencarian *centroid* baru dapat dilihat di Tabel 3.20.

Tabel 3.20 Hasil *centroid* baru tiga *cluster* di iterasi ketiga

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	11.92481	13.26139	0.85133	5.22453	2.86747	4.57543	5.06119
2	18.75833	16.31333	0.88513	6.21450	3.72693	3.57145	6.06920
3	14.79465	14.52113	0.88094	5.58287	3.29804	2.83135	5.20658

Tabel 3.21 Hasil perhitungan di iterasi ketiga untuk tiga *cluster*

Cluster	Hasil
1	4.43531
2	4.16295
3	0.85070

Langkah selanjutnya mengelompokkan kembali dengan persamaan (2,2) menggunakan *centroid* baru. Hasil perhitungan data pertama di iterasi ketiga menggunakan *centroid* baru dapat dilihat di Tabel 3.21.

Tabel 3.22 Hasil mengelompokkan tiga *cluster* di iterasi ketiga.

NO	C1	C2	C3	TERDEKAT	CLUSTER
1	4.43531	4.16295	0.85070	0.85070	C3
2	4.84129	5.14115	1.83379	1.83379	C3
3	3.17573	5.31260	0.83123	0.83123	C3
4	3.13583	5.83589	1.34599	1.34599	C3
5	5.64030	3.82836	2.07033	2.07033	C3
6	3.41043	5.18870	0.71211	0.71211	C3
7	3.22929	4.60894	0.76386	0.76386	C3
8	3.03775	5.40558	0.85596	0.85596	C3
9	5.92542	2.78120	2.36107	2.36107	C3
10	5.67453	3.08532	2.05480	2.05480	C3
11	3.75397	4.04720	1.81285	1.81285	C3
12	3.68391	5.69681	1.42462	1.42462	C3
13	2.24646	5.63774	1.62876	1.62876	C3
14	2.51711	5.68781	1.21568	1.21568	C3
15	2.60789	5.75983	1.23857	1.23857	C3
16	2.93373	4.94433	1.47269	1.47269	C3
17	2.31990	5.88617	2.70236	2.31990	C1
18	5.07587	4.16108	1.56490	1.56490	C3
19	4.12570	5.22065	1.31213	1.31213	C3
20	1.00330	6.85859	2.66430	1.00330	C1
21	2.97024	5.14983	0.71486	0.71486	C3
22	3.08830	5.30296	0.76064	0.76064	C3
23	5.78050	4.41805	2.37687	2.37687	C3
24	3.16924	7.86484	3.38646	3.16924	C1
25	4.47015	4.60300	1.12853	1.12853	C3

Hasil akhir perhitungan *euclidean distance* di iterasi ketiga dapat dilihat di Tabel 3.22 yang menampilkan contoh hasil akhir 25 data pertama. Penentuan suatu data masuk ke *cluster* pertama atau kedua atau ketiga adalah dengan mencari hasil akhir perhitungan *euclidean distance* yang terdekat antar *cluster*. Sehingga dari iterasi ketiga didapatkan jumlah anggota *cluster* pertama 82 data, *cluster* kedua 61 data, dan *cluster* ketiga 67 data. Dikarenakan anggota *cluster* sebelumnya dengan anggota *cluster* baru tidak sama atau berubah, maka proses *clustering* akan dilanjutkan

dengan mengulangi kembali proses *clustering* dalam pencarian *centroid* baru menggunakan anggota *cluster* terakhir.

Langkah mencari *centroid* baru dengan menggunakan persamaan (2,1). Perhitungan *cluster* pertama dapat dilihat sebagai berikut :

$$V_{\text{Area A}} = \frac{1734.53000}{82} = 11.98866$$

$$V_{\text{Perimeter P}} = \frac{1089.32000}{82} = 13.28439$$

$$V_{\text{Compactness C}} = \frac{69.92440}{82} = 0.85274$$

$$V_{\text{Length of Kernel}} = \frac{428.64900}{82} = 5.22743$$

$$V_{\text{Width of Kernel}} = \frac{405.42300}{134} = 2.88009$$

$$V_{\text{Asymmetry Coefficient}} = \frac{375.88200}{82} = 4.58393$$

$$V_{\text{Length of Kernel Groove}} = \frac{414.48500}{82} = 5.05470$$

Perhitungan *centroid cluster* kedua dan ketiga juga menggunakan persamaan (2,1) dan hasil akhir perhitungan pencarian *centroid* baru dapat dilihat di Tabel 3.23.

Tabel 3.23 Hasil *centroid* baru tiga *cluster* di iterasi keempat

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	11.98866	13.28439	0.85274	5.22743	2.88009	4.58393	5.05470
2	18.72180	16.29738	0.88509	6.20893	3.72267	3.60359	6.06610
3	14.81910	14.53716	0.88052	5.59101	3.29936	2.70256	5.21099

Langkah selanjutnya mengelompokkan kembali dengan persamaan (2,2) menggunakan *centroid* baru. Hasil perhitungan data pertama di iterasi keempat menggunakan *centroid* baru dapat dilihat Tabel 3.24. Hasil akhir perhitungan *euclidean distance* di iterasi ketiga dapat dilihat di Tabel 3.24 yang menampilkan contoh hasil akhir 25 data pertama.

Tabel 3.24 Hasil perhitungan di iterasi keempat untuk tiga *cluster*

Cluster	Hasil
1	4.38247
2	4.13560
3	0.74021

Tabel 3.25 Hasil mengelompokkan tiga *cluster* di iterasi keempat.

NO	C1	C2	C3	TERDEKAT	CLUSTER
1	4.38247	4.13560	0.74021	0.74021	C3
2	4.80113	5.12266	1.70588	1.70588	C3
3	3.12513	5.27861	0.84908	0.84908	C3
4	3.09596	5.80421	1.32464	1.32464	C3
5	5.58893	3.81508	1.95985	1.95985	C3
6	3.36171	5.15654	0.68201	0.68201	C3
7	3.16710	4.56861	0.89546	0.89546	C3
8	2.98899	5.37119	0.87626	0.87626	C3
9	5.86918	2.76554	2.29225	2.29225	C3
10	5.61850	3.06801	1.97621	1.97621	C3
11	3.68643	4.00042	1.92578	1.92578	C3
12	3.64746	5.66935	1.34664	1.34664	C3
13	2.18214	5.59552	1.74209	1.74209	C3
14	2.46583	5.65029	1.28192	1.28192	C3
15	2.56027	5.72359	1.28441	1.28441	C3
16	2.86615	4.90094	1.60008	1.60008	C3
17	2.25156	5.83884	2.83052	2.25156	C1
18	5.02516	4.14176	1.44901	1.44901	C3
19	4.08112	5.19499	1.22078	1.22078	C3
20	0.94724	6.81588	2.75397	0.94724	C1
21	2.91673	5.11289	0.79053	0.79053	C3
22	3.04012	5.26858	0.77608	0.77608	C3
23	5.73456	4.40820	2.25155	2.25155	C3
24	3.17521	7.83484	3.36244	3.17521	C1
25	4.42244	4.57884	1.00175	1.00175	C3

Penentuan suatu data masuk ke *cluster* pertama atau kedua atau ketiga adalah dengan mencari hasil akhir perhitungan *euclidean distance* yang terdekat antar *cluster*. Sehingga dari iterasi keempat didapatkan jumlah anggota *cluster* pertama 82 data, *cluster* kedua 61 data, dan *cluster* ketiga 67 data. Proses *clustering* berhenti di iterasi keempat karena anggota

cluster sebelumnya dengan anggota *cluster* baru sama atau tidak mengalami perubahan. Sehingga proses *clustering* tiga *cluster* dengan penentuan *centroid* awal menggunakan penentuan metode ROC menghasilkan *cluster* pertama dengan anggota 82, *cluster* kedua dengan anggota 61, dan *cluster* ketiga dengan anggota 67.

3.3.3 Empat Cluster

Nilai *centroid* dari perhitungan ROC yang diambil sebagai *centroid* awal untuk empat *cluster* adalah nilai yang memiliki nilai tertinggi, nilai terendah, dan nilai tengah. Nilai yang terpilih untuk *centroid* awal adalah data dengan nilai tertinggi di id 61, untuk data dengan nilai terendah di id 90, untuk data dengan nilai tertinggi kedua di id 178, dan untuk data dengan nilai terendah di id 89 atau bisa dilihat pada Tabel 3.26.

Tabel 3.26 *Centroid* awal empat *cluster* dengan ROC.

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	11.42000	12.86000	0.86830	5.00800	2.85000	2.70000	4.00000
2	20.88000	17.05000	0.90310	6.45000	4.03200	5.01600	6.32100
3	10.74000	12.73000	0.83290	5.14500	2.64200	4.70200	4.96300
4	21.18000	17.21000	0.89890	6.57300	4.03300	5.78000	6.23100

Tabel 3.27 Hasil perhitungan di iterasi pertama untuk empat *cluster*.

Cluster	Hasil
1	4.60080
2	6.81793
3	5.65121
4	7.45175

Langkah selanjutnya mencari *euclidean distance* dari setiap data dengan *centroid* awal yang telah ditentukan menggunakan persamaan (2,2) dan menggunakan *centroid* yang di Tabel 3.26. Hasil perhitungan data pertama di iterasi pertama menggunakan *euclidean distance* dapat dilihat di Tabel 3.27.

Tabel 3.28 Hasil mengelompokkan empat *cluster* di iterasi pertama

NO	C1	C2	C3	C4	TERDEKAT	CLUSTER
1	4.60080	6.81793	5.65121	7.45175	4.60080	C1
2	4.37839	7.82876	5.89442	8.51414	4.37839	C1
3	3.27857	7.85000	4.35806	8.41450	3.27857	C1
4	2.87154	8.41590	4.20077	8.99929	2.87154	C1
5	5.56198	6.50197	6.82858	7.21698	5.56198	C1
6	3.45134	7.76867	4.57914	8.35485	3.45134	C1
7	4.01201	7.03661	4.53527	7.54823	4.01201	C1
8	3.18563	7.95225	4.21400	8.51602	3.18563	C1
9	6.27161	5.48855	7.18124	6.20254	5.48855	C2
10	5.91555	5.78421	6.92304	6.48643	5.78421	C2
11	4.94783	6.23130	5.07527	6.65969	4.94783	C1
12	3.28308	8.34360	4.70969	8.96809	3.28308	C1
13	3.15482	7.95764	3.54254	8.41126	3.15482	C1
14	2.87674	8.16670	3.72139	8.68830	2.87674	C1
15	2.79774	8.26688	3.77259	8.80128	2.79774	C1
16	3.90287	7.21080	4.24850	7.65786	3.90287	C1
17	3.85692	7.90270	3.55608	8.24973	3.55608	C3
18	4.98227	6.82992	6.25540	7.51267	4.98227	C1
19	3.78062	7.84093	5.22247	8.47293	3.78062	C1
20	2.25438	9.16388	2.27198	9.59974	2.25438	C1
21	3.44978	7.66262	4.20578	8.20237	3.44978	C1
22	3.32348	7.86533	4.27064	8.43385	3.32348	C1
23	5.45820	7.08332	6.90075	7.81432	5.45820	C1
24	1.77846	10.48222	3.59717	11.07672	1.77846	C1
25	4.36948	7.27336	5.62215	7.92042	4.36948	C1

Hasil pengelompokkan menggunakan hasil perhitungan *euclidean distance* di iterasi pertama dapat dilihat di tabel 3.28 yang menampilkan contoh hasil akhir 25 data pertama. Penentuan suatu data masuk ke *cluster* pertama atau kedua atau ketiga atau keempat adalah dengan mencari hasil akhir perhitungan *euclidean distance* yang terdekat antar *cluster*. Sehingga dari iterasi pertama didapatkan jumlah anggota *cluster* pertama 83 data, *cluster* kedua 65 data, *cluster* ketiga 59 data, dan *cluster* keempat 3 data.

Langkah selanjutnya mencari *centroid* baru untuk iterasi kedua dengan membagi jumlah semua data tiap *cluster* dengan jumlah anggota *cluster* di tiap atribut *dataset* atau menggunakan persamaan (2,1). Perhitungan *cluster* pertama dapat dilihat sebagai berikut :

$$V_{Area A} = \frac{1152.90000}{83} = 13.89036$$

$$V_{\text{Perimeter P}} = \frac{1170.80000}{83} = 14.10602$$

$$V_{\text{Compactness C}} = \frac{72.63160}{83} = 0.87508$$

$$V_{\text{Length of Kernel}} = \frac{452.53500}{83} = 5.45223$$

$$V_{\text{Width of Kernel}} = \frac{263.37700}{83} = 3.17322$$

$$V_{\text{Asymmetry Coefficient}} = \frac{221.83220}{83} = 2.67268$$

$$V_{\text{Length of Kernel Groove}} = \frac{421.35200}{83} = 5.07653$$

Hasil perhitungan *centroid cluster* kedua, ketiga, dan keempat juga menggunakan persamaan (2,1) dan hasil akhir perhitungan pencarian *centroid* baru dapat dilihat di Tabel 3.29.

Tabel 3.29 Hasil *centroid* baru empat *cluster* di iterasi kedua

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	13.89036	14.10602	0.87508	5.45223	3.17322	2.67268	5.07653
2	18.39554	16.16200	0.88401	6.16243	3.68560	3.45886	6.02362
3	12.01373	13.31169	0.85012	5.25003	2.87515	5.28392	5.12363
4	20.18667	16.91000	0.88693	6.38233	3.91067	6.12100	6.15733

Tabel 3.30 Hasil perhitungan di iterasi kedua untuk empat *cluster*

Cluster	Hasil
1	1.65984
2	3.74922
3	4.76646
4	6.73704

Langkah selanjutnya mengelompokkan kembali dengan persamaan (2,2) menggunakan *centroid* baru. Sebagai contoh perhitungan data pertama di iterasi kedua menggunakan *centroid* baru dapat dilihat di Tabel 3.30. Hasil akhir perhitungan *euclidean distance* di iterasi kedua dapat dilihat di Tabel 3.31 yang menampilkan contoh hasil akhir 25 data pertama.

Tabel 3.31 Hasil mengelompokkan empat *cluster* di iterasi kedua

NO	C1	C2	C3	C4	TERDEKAT	CLUSTER
1	1.65984	3.74922	4.76646	6.73704	1.65984	C1
2	1.99576	4.74190	5.32240	7.88289	1.99576	C1
3	0.52692	4.90347	3.57437	7.59794	0.52692	C1
4	0.57807	5.42328	3.63910	8.20047	0.57807	C1
5	2.78972	3.44638	5.99357	6.66244	2.78972	C1
6	0.57751	4.77608	3.82215	7.56241	0.57751	C1
7	1.28879	4.21088	3.41892	6.66877	1.28879	C1
8	0.26974	4.99397	3.45260	7.69163	0.26974	C1
9	3.29028	2.38858	6.16514	5.63851	2.38858	C2
10	2.96809	2.68820	5.94182	5.90211	2.68820	C2
11	2.46114	3.69187	3.72023	5.71061	2.46114	C1
12	0.97079	5.28256	4.20301	8.22161	0.97079	C1
13	1.35935	5.25156	2.44896	7.48126	1.35935	C1
14	0.52136	5.28201	2.91350	7.81766	0.52136	C1
15	0.40082	5.35151	3.04257	7.94540	0.40082	C1
16	1.71190	4.56929	3.02027	6.72900	1.71190	C1
17	2.62502	5.54290	2.13867	7.24166	2.13867	C3
18	2.22031	3.76327	5.44576	6.88901	2.22031	C1
19	1.34827	4.81510	4.57956	7.75955	1.34827	C1
20	1.94759	6.46974	1.42743	8.63764	1.42743	C3
21	0.61078	4.74104	3.30411	7.34961	0.61078	C1
22	0.31163	4.88936	3.49294	7.60984	0.31163	C1
23	2.89268	4.04663	6.19907	7.29092	2.89268	C1
24	2.41268	7.44864	3.87722	10.26818	2.41268	C1
25	1.60717	4.19391	4.87014	7.23201	1.60717	C1

Penentuan suatu data masuk ke *cluster* pertama atau kedua atau ketiga atau keempat adalah dengan mencari hasil akhir perhitungan *euclidean distance* yang terdekat antar *cluster*. Sehingga dari iterasi kedua didapatkan jumlah anggota *cluster* pertama 76 data, *cluster* kedua 58 data, *cluster* ketiga 66 data, dan *cluster* keempat 10 data. Dikarenakan anggota *cluster* sebelumnya dengan anggota *cluster* baru tidak sama atau berubah, maka proses *clustering* akan dilanjutkan dengan mengulangi kembali proses *clustering* dalam pencarian *centroid* baru menggunakan anggota *cluster* terakhir.

Langkah mencari *centroid* baru dengan menggunakan persamaan (2,1). Perhitungan *cluster* pertama dapat dilihat sebagai berikut :

$$V_{Area A} = \frac{1077.16000}{76} = 14.17316$$

$$V_{\text{Perimeter P}} = \frac{1081.9100}{76} = 14.23566$$

$$V_{\text{Compactness C}} = \frac{66.6722}{76} = 0.87727$$

$$V_{\text{Length of Kernel}} = \frac{417.413}{76} = 5.49228$$

$$V_{\text{Width of Kernel}} = \frac{244.211}{76} = 3.21330$$

$$V_{\text{Asymmetry Coefficient}} = \frac{200.5332}{76} = 2.63859$$

$$V_{\text{Length of Kernel Groove}} = \frac{388.03200}{76} = 5.10568$$

Perhitungan *centroid cluster* kedua, ketiga dan keempat juga menggunakan persamaan (2,1) dan hasil akhir perhitungan pencarian *centroid* baru dapat dilihat di Tabel 3.32.

Tabel 3.32 Hasil *centroid* baru empat *cluster* di iterasi ketiga

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	14.17316	14.23566	0.87727	5.49228	3.21330	2.63859	5.10568
2	18.21897	16.09276	0.88321	6.14126	3.66391	3.26069	6.00245
3	11.88712	13.24667	0.85025	5.22536	2.86061	5.04621	5.08506
4	19.95700	16.78800	0.88948	6.35120	3.87890	5.40690	6.18650

Tabel 3.33 Hasil perhitungan di iterasi ketiga untuk empat *cluster*

Cluster	Hasil
1	1.34795
2	3.50502
3	4.73374
4	6.13256

Langkah selanjutnya mengelompokkan dengan persamaan (2,2) menggunakan *centroid* baru. Hasil perhitungan data pertama di iterasi ketiga menggunakan *centroid* baru dapat dilihat di Tabel 3.33.

Tabel 3.34 Hasil mengelompokkan empat *cluster* di iterasi ketiga

NO	C1	C2	C3	C4	TERDEKAT	CLUSTER
1	1.34795	3.50502	4.73374	6.13256	1.34795	C1
2	1.81061	4.47734	5.22339	7.23917	1.81061	C1
3	0.41699	4.68828	3.50661	7.07096	0.41699	C1
4	0.69930	5.19579	3.52448	7.65844	0.69930	C1
5	2.49793	3.17787	5.95326	5.98535	2.49793	C1
6	0.34407	4.55080	3.75084	7.01346	0.34407	C1
7	1.11773	4.02994	3.43886	6.18704	1.11773	C1
8	0.22498	4.77757	3.37871	7.16574	0.22498	C1
9	2.97760	2.11623	6.17326	4.95225	2.11623	C2
10	2.65275	2.41977	5.93822	5.22867	2.41977	C2
11	2.29757	3.57477	3.82586	5.80446	2.29757	C1
12	0.94319	5.03763	4.08397	7.63872	0.94319	C1
13	1.44241	5.08631	2.45268	7.06270	1.44241	C1
14	0.70065	5.08199	2.84448	7.33251	0.70065	C1
15	0.63003	5.14447	2.95748	7.44764	0.63003	C1
16	1.64538	4.41996	3.07389	6.31361	1.64538	C1
17	2.68505	5.43280	2.27573	6.93724	2.27573	C3
18	1.93445	3.50357	5.39701	6.24294	1.93445	C1
19	1.18080	4.57248	4.48941	7.16797	1.18080	C1
20	2.19796	6.30171	1.32471	8.24663	1.32471	C3
21	0.50487	4.53647	3.26515	6.84284	0.50487	C1
22	0.15070	4.67080	3.42475	7.07836	0.15070	C1
23	2.63958	3.77502	6.13016	6.60370	2.63958	C1
24	2.67254	7.21040	3.64149	9.72632	2.67254	C1
25	1.33933	3.94163	4.80746	6.61696	1.33933	C1

Hasil akhir perhitungan *euclidean distance* di iterasi ketiga dapat dilihat di Tabel 3.34 yang menampilkan contoh hasil akhir 25 data pertama. Penentuan suatu data masuk ke *cluster* pertama atau kedua atau ketiga atau keempat adalah dengan mencari hasil akhir perhitungan *euclidean distance* yang terdekat antar *cluster*. Sehingga dari iterasi ketiga didapatkan jumlah anggota *cluster* pertama 72 data, *cluster* kedua 56 data, *cluster* ketiga 70 data, *cluster* keempat 12 data. Dikarenakan anggota *cluster* sebelumnya dengan anggota *cluster* baru tidak sama atau berubah, maka proses *clustering* akan dilanjutkan dengan mengulangi kembali proses *clustering* dalam pencarian *centroid* baru menggunakan anggota *cluster* terakhir.

Langkah selanjutnya mencari *centroid* baru dengan membagi jumlah semua data tiap *cluster* dengan jumlah anggota *cluster* di tiap atribut

dataset atau menggunakan persamaan (2,1). Perhitungan *cluster* pertama dapat dilihat sebagai berikut :

$$V_{Area A} = \frac{1028.92000}{72} = 14.29056$$

$$V_{Perimeter P} = \frac{1029.12000}{72} = 14.29333$$

$$V_{Compactness C} = \frac{63.19440}{72} = 0.87770$$

$$V_{Length\ of\ Kernel} = \frac{396.8640}{72} = 5.51200$$

$$V_{Width\ of\ Kernel} = \frac{232.50100}{72} = 3.22918$$

$$V_{Asymmetry\ Coefficient} = \frac{188.66620}{72} = 2.62036$$

$$V_{Length\ of\ Kernel\ Groove} = \frac{368.50400}{72} = 5.11811$$

Perhitungan *centroid cluster* kedua, ketiga dan keempat juga menggunakan persamaan (2,1) dan hasil akhir perhitungan pencarian *centroid* baru dapat dilihat di Tabel 3.35.

Tabel 3.35 Hasil *centroid* baru empat *cluster* di iterasi keempat

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	14.29056	14.29333	0.87770	5.51200	3.22918	2.62036	5.11811
2	18.19429	16.07786	0.88362	6.13239	3.66304	3.21213	5.99438
3	11.89700	13.24386	0.85134	5.22033	2.86443	4.92739	5.07346
4	19.78250	16.74167	0.88654	6.35758	3.84717	5.27583	6.19350

Tabel 3.36 Hasil perhitungan di iterasi keempat untuk empat *cluster*

Cluster	Hasil
1	1.21591
2	3.46181
3	4.65818
4	5.91517

Langkah selanjutnya mengelompokkan kembali dengan persamaan (2,2) menggunakan *centroid* baru. Hasil perhitungan data pertama di iterasi keempat menggunakan *centroid* baru dapat dilihat di tabel 3.36

Tabel 3.37 Hasil mengelompokkan empat *cluster* di iterasi keempat

NO	C1	C2	C3	C4	TERDEKAT	CLUSTER
1	1.21591	3.46181	4.65818	5.91517	1.21591	C1
2	1.74080	4.42651	5.12697	7.02292	1.74080	C1
3	0.44120	4.65193	3.42068	6.86367	0.44120	C1
4	0.78395	5.15632	3.42486	7.44845	0.78395	C1
5	2.37548	3.12401	5.87411	5.77213	2.37548	C1
6	0.29890	4.51210	3.66338	6.80264	0.29890	C1
7	1.06975	4.00372	3.38351	5.98215	1.06975	C1
8	0.32283	4.74155	3.29088	6.95639	0.32283	C1
9	2.84520	2.06446	6.11145	4.73188	2.06446	C2
10	2.51919	2.36808	5.87187	5.01100	2.36808	C2
11	2.24179	3.56431	3.80548	5.11104	2.24179	C1
12	0.96006	4.99373	3.98275	7.42340	0.96006	C1
13	1.50032	5.06312	2.39464	6.86841	1.50032	C1
14	0.80499	5.05047	2.75934	7.12715	0.80499	C1
15	0.74802	5.11110	2.86725	7.24069	0.74802	C1
16	1.63974	4.40019	3.03339	6.12247	1.63974	C1
17	2.72571	5.42178	2.27759	6.76597	2.27759	C3
18	1.81590	3.45332	5.31537	6.02929	1.81590	C1
19	1.12900	4.52739	4.39598	6.95795	1.12900	C1
20	2.31108	6.27814	1.23482	8.05427	1.23482	C3
21	0.51718	4.50454	3.18942	6.63280	0.51718	C1
22	0.22870	4.63488	3.33891	6.86626	0.22870	C1
23	2.53495	3.71974	6.04196	6.39111	2.53495	C1
24	2.78415	7.16862	3.52182	9.51315	2.78415	C1
25	1.23047	3.89561	4.72229	6.39970	1.23047	C1

Hasil akhir perhitungan *euclidean distance* di iterasi ketiga dapat dilihat di Tabel 3.37 yang menampilkan contoh hasil akhir 25 data pertama. Penentuan suatu data masuk ke *cluster* pertama atau kedua atau ketiga atau keempat adalah dengan mencari hasil akhir perhitungan *euclidean distance* yang terdekat antar *cluster*. Sehingga dari iterasi keempat didapatkan jumlah anggota *cluster* pertama 71 data, *cluster* kedua 55 data, *cluster* ketiga 71 data, dan *cluster* keempat 12 data. Dikarenakan anggota *cluster* sebelumnya dengan anggota *cluster* baru tidak sama atau berubah, maka proses *clustering* akan dilanjutkan dengan mengulangi kembali proses

clustering dalam pencarian *centroid* baru menggunakan anggota *cluster* terakhir.

Langkah selanjutnya mencari *centroid* baru dengan membagi jumlah semua data tiap *cluster* dengan jumlah anggota *cluster* di tiap atribut *dataset* atau menggunakan persamaan (2,1). Perhitungan *cluster* pertama dapat dilihat sebagai berikut :

$$V_{Area A} = \frac{1019.77000}{71} = 14.36296$$

$$V_{Perimeter P} = \frac{1016.97000}{71} = 14.32352$$

$$V_{Compactness C} = \frac{62.37390}{71} = 0.87851$$

$$V_{Length\ of\ Kernel} = \frac{391.86800}{71} = 5.51927$$

$$V_{Width\ of\ Kernel} = \frac{230.17200}{71} = 3.24186$$

$$V_{Asymmetry\ Coefficient} = \frac{184.65020}{71} = 2.60071$$

$$V_{Length\ of\ Kernel\ Groove} = \frac{363.73300}{71} = 5.12300$$

Perhitungan *centroid cluster* kedua, ketiga dan keempat juga menggunakan persamaan (2,1) dan hasil akhir perhitungan pencarian *centroid* baru dapat dilihat di Tabel 3.38.

Tabel 3.38 Hasil *centroid* baru empat *cluster* di iterasi kelima

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	14.36296	14.32352	0.87851	5.51927	3.24186	2.60071	5.12300
2	18.23055	16.09255	0.88381	6.13796	3.66665	3.21920	6.00287
3	11.91861	13.25681	0.85122	5.22563	2.86532	4.88550	5.07556
4	19.78250	16.74167	0.88654	6.35758	3.84717	5.27583	6.19350

Langkah selanjutnya mengelompokkan dengan persamaan (2,2) menggunakan *centroid* baru. Sebagai contoh perhitungan data pertama di iterasi kelima menggunakan *centroid* baru dapat dilihat di Tabel 3.39.

Tabel 3.39 Hasil perhitungan di iterasi kelima untuk empat *cluster*

Cluster	Hasil
1	1.13552
2	3.50269
3	4.61308
4	5.91517

Hasil akhir perhitungan *euclidean distance* di iterasi keempat dapat dilihat di Tabel 3.39 yang menampilkan contoh hasil akhir 25 data pertama. Penentuan suatu data masuk ke *cluster* pertama atau kedua atau ketiga atau keempat adalah dengan mencari hasil akhir perhitungan *euclidean distance* yang terdekat antar *cluster*.

Tabel 3.40 Hasil mengelompokkan empat *cluster* di iterasi kelima

NO	C1	C2	C3	C4	TERDEKAT	CLUSTER
1	1.13552	3.50269	4.61308	5.91517	1.13552	C1
2	1.69424	4.46518	5.07874	7.02292	1.69424	C1
3	0.46923	4.69282	3.37504	6.86367	0.46923	C1
4	0.83408	5.19737	3.37723	7.44845	0.83408	C1
5	2.29759	3.16043	5.82867	5.77213	2.29759	C1
6	0.28855	4.55317	3.61688	6.80264	0.28855	C1
7	1.05683	4.04344	3.34369	5.98215	1.05683	C1
8	0.39060	4.78257	3.24428	6.95639	0.39060	C1
9	2.76562	2.10147	6.06920	4.73188	2.10147	C2
10	2.43762	2.40632	5.82881	5.01100	2.40632	C2
11	2.22010	3.59935	3.77598	5.11104	2.22010	C1
12	0.97041	5.03453	3.93412	7.42340	0.97041	C1
13	1.54643	5.10205	2.35587	6.86841	1.54643	C1
14	0.87625	5.09112	2.71319	7.12715	0.87625	C1
15	0.82414	5.15196	2.82022	7.24069	0.82414	C1
16	1.64847	4.43796	2.99958	6.12247	1.64847	C1
17	2.75963	5.45653	2.26081	6.76597	2.26081	C3
18	1.73987	3.49206	5.26968	6.02929	1.73987	C1
19	1.09389	4.56752	4.34937	6.95795	1.09389	C1
20	2.38495	6.31726	1.18926	8.05427	1.18926	C3
21	0.55168	4.54527	3.14422	6.63280	0.55168	C1
22	0.30059	4.67594	3.29194	6.86626	0.30059	C1
23	2.46671	3.75465	5.99488	6.39111	2.46671	C1
24	2.84783	7.20955	3.47930	9.51315	2.84783	C1
25	1.16302	3.93586	4.67542	6.39970	1.16302	C1

Sehingga dari iterasi kelima didapatkan jumlah anggota *cluster* pertama 69 data, *cluster* kedua 55 data, *cluster* ketiga 74 data, dan *cluster* keempat 12 data. Dikarenakan anggota *cluster* sebelumnya dengan anggota *cluster* baru tidak sama atau berubah, maka proses *clustering* akan dilanjutkan dengan mengulangi kembali proses *clustering* dalam pencarian *centroid* baru menggunakan anggota *cluster* terakhir.

Langkah mencari *centroid* baru dengan menggunakan persamaan (2,1). Perhitungan *cluster* pertama dapat dilihat sebagai berikut :

$$V_{Area A} = \frac{994.53000}{69} = 14.41348$$

$$V_{Perimeter P} = \frac{989.59000}{69} = 14.34188$$

$$V_{Compactness C} = \frac{60.68230}{69} = 0.87945$$

$$V_{Length\ of\ Kernel} = \frac{381.03100}{69} = 5.52219$$

$$V_{Width\ of\ Kernel} = \frac{224.38200}{69} = 3.25191$$

$$V_{Asymmetry\ Coefficient} = \frac{178.30820}{69} = 2.58418$$

$$V_{Length\ of\ Kernel\ Groove} = \frac{352.92600}{69} = 5.11487$$

Perhitungan *centroid cluster* kedua, ketiga dan keempat juga menggunakan persamaan (2,1) dan hasil akhir perhitungan pencarian *centroid* baru dapat dilihat di Tabel 3.41.

Tabel 3.41 Hasil *centroid* baru empat *cluster* di iterasi keenam

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	14.41348	14.34188	0.87945	5.52219	3.25191	2.58418	5.11487
2	18.23055	16.09255	0.88381	6.13796	3.66665	3.21920	6.00287
3	11.93757	13.26851	0.85108	5.23084	2.86612	4.83916	5.08442
4	19.78250	16.74167	0.88654	6.35758	3.84717	5.27583	6.19350

Langkah selanjutnya mengelompokkan kembali dengan persamaan (2,2) menggunakan *centroid* baru. Hasil perhitungan data pertama di iterasi keenam menggunakan *centroid* baru dapat dilihat Tabel 3.42.

Tabel 3.42 Hasil perhitungan di iterasi keenam untuk empat *cluster*

Cluster	Hasil
1	1.08136
2	3.50269
3	4.56768
4	5.91517

Tabel 3.43 Hasil mengelompokkan empat *cluster* di iterasi keenam

NO	C1	C2	C3	C4	TERDEKAT	CLUSTER
1	1.08136	3.50269	4.56768	5.91517	1.08136	C1
2	1.65995	4.46518	5.02920	7.02292	1.65995	C1
3	0.48707	4.69282	3.32936	6.86367	0.48707	C1
4	0.86477	5.19737	3.32854	7.44845	0.86477	C1
5	2.24290	3.16043	5.78282	5.77213	2.24290	C1
6	0.28475	4.55317	3.56989	6.80264	0.28475	C1
7	1.05575	4.04344	3.30484	5.98215	1.05575	C1
8	0.43624	4.78257	3.19708	6.95639	0.43624	C1
9	2.71428	2.10147	6.02662	4.73188	2.10147	C2
10	2.38323	2.40632	5.78562	5.01100	2.38323	C1
11	2.21097	3.59935	3.74902	5.11104	2.21097	C1
12	0.97713	5.03453	3.88378	7.42340	0.97713	C1
13	1.57942	5.10205	2.31951	6.86841	1.57942	C1
14	0.92484	5.09112	2.66704	7.12715	0.92484	C1
15	0.87432	5.15196	2.77293	7.24069	0.87432	C1
16	1.65692	4.43796	2.96871	6.12247	1.65692	C1
17	2.78490	5.45653	2.24942	6.76597	2.24942	C3
18	1.68546	3.49206	5.22361	6.02929	1.68546	C1
19	1.06397	4.56752	4.30231	6.95795	1.06397	C1
20	2.43656	6.31726	1.14426	8.05427	1.14426	C3
21	0.58595	4.54527	3.09876	6.63280	0.58595	C1
22	0.35711	4.67594	3.24397	6.86626	0.35711	C1
23	2.41794	3.75465	5.94689	6.39111	2.41794	C1
24	2.88983	7.20955	3.43284	9.51315	2.88983	C1
25	1.11543	3.93586	4.62795	6.39970	1.11543	C1

Hasil akhir perhitungan *euclidean distance* di iterasi keempat dapat dilihat di tabel 3.43 yang menampilkan contoh hasil akhir 25 data pertama. Penentuan suatu data masuk ke *cluster* pertama atau kedua atau ketiga atau keempat atau keempat adalah dengan mencari hasil akhir perhitungan *euclidean distance* yang terdekat antar *cluster*.

Sehingga dari iterasi keenam didapatkan jumlah anggota *cluster* pertama 69 data, *cluster* kedua 54 data, *cluster* ketiga 74 data dan *cluster* keempat 12 data. Dikarenakan anggota *cluster* sebelumnya dengan anggota *cluster* baru tidak sama atau berubah, maka proses *clustering* akan dilanjutkan dengan mengulangi kembali proses *clustering* dalam pencarian *centroid* baru menggunakan anggota *cluster* terakhir.

Langkah mencari *centroid* baru dengan menggunakan persamaan (2.1). Perhitungan *cluster* pertama dapat dilihat sebagai berikut :

$$V_{\text{Area A}} = \frac{998.87000}{69} = 14.47638$$

$$V_{\text{Perimeter P}} = \frac{991.69000}{69} = 14.37232$$

$$V_{\text{Compactness C}} = \frac{60.69100}{69} = 0.87958$$

$$V_{\text{Length of Kernel}} = \frac{381.81000}{69} = 5.53348$$

$$V_{\text{Width of Kernel}} = \frac{224.94600}{69} = 3.26009$$

$$V_{\text{Asymmetry Coefficient}} = \frac{178.07620}{69} = 2.58081$$

$$V_{\text{Length of Kernel Groove}} = \frac{353.40300}{69} = 5.12178$$

Tabel 3.44 Hasil *centroid* baru empat *cluster* di iterasi ketujuh

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	14.47638	14.37232	0.87958	5.53348	3.26009	2.58081	5.12178
2	18.26370	16.10815	0.88373	6.14267	3.66965	3.24235	6.01157
3	11.93973	13.26693	0.85145	5.22916	2.86712	4.80399	5.08404
4	19.78250	16.74167	0.88654	6.35758	3.84717	5.27583	6.19350

Tabel 3.45 Hasil perhitungan di iterasi ketujuh untuk empat *cluster*

Cluster	Hasil
1	1.01358
2	3.54576
3	4.54668
4	5.91517

Perhitungan *centroid cluster* kedua, ketiga dan keempat juga menggunakan persamaan (2,1) dan hasil akhir perhitungan pencarian *centroid* baru dapat dilihat di Tabel 3.44.

Langkah selanjutnya mengelompokkan kembali dengan persamaan (2,2) menggunakan *centroid* baru. Sebagai contoh perhitungan data pertama di iterasi ketujuh menggunakan *centroid* baru dapat dilihat di Tabel 3.45. Hasil akhir perhitungan *euclidean distance* di iterasi keempat dapat dilihat di tabel 3.46 yang menampilkan contoh hasil akhir 25 data pertama.

Penentuan suatu data masuk ke *cluster* pertama atau kedua atau ketiga atau keempat adalah dengan mencari hasil akhir perhitungan *euclidean distance* yang terdekat antar *cluster*. Sehingga dari iterasi keempat didapatkan jumlah anggota *cluster* pertama 69 data, *cluster* kedua 54 data, dan *cluster* ketiga 75 data, dan *cluster* keempat 12 data. Proses *clustering* berhenti di iterasi keempat karena anggota *cluster* sebelumnya dengan anggota *cluster* baru sama atau tidak mengalami perubahan. Sehingga proses *clustering* empat *cluster* dengan penentuan *centroid* awal menggunakan penentuan metode ROC menghasilkan *cluster* pertama

dengan anggota 69 data, *cluster* kedua dengan anggota 54 data, dan *cluster* ketiga dengan anggota 75 data, dan *cluster* keempat dengan anggota 12 data.

Tabel 3.46 Hasil mengelompokkan empat *cluster* di iterasi ketujuh

NO	C1	C2	C3	C4	TERDEKAT	CLUSTER
1	1.01358	3.54576	4.54668	5.91517	1.01358	C1
2	1.63634	4.50968	5.00167	7.02292	1.63634	C1
3	0.53089	4.73317	3.30556	6.86367	0.53089	C1
4	0.92416	5.23905	3.30027	7.44845	0.92416	C1
5	2.18217	3.20456	5.76059	5.77213	2.18217	C1
6	0.31925	4.59455	3.54542	6.80264	0.31925	C1
7	1.03933	4.07930	3.29050	5.98215	1.03933	C1
8	0.50219	4.82295	3.17254	6.95639	0.50219	C1
9	2.64498	2.14542	6.00943	4.73188	2.14542	C2
10	2.31354	2.45088	5.76721	5.01100	2.31354	C1
11	2.18229	3.62634	3.74529	5.11104	2.18229	C1
12	1.00881	5.07776	3.85474	7.42340	1.00881	C1
13	1.61324	5.13628	2.30543	6.86841	1.61324	C1
14	0.98332	5.12960	2.64370	7.12715	0.98332	C1
15	0.93892	5.19129	2.74791	7.24069	0.93892	C1
16	1.65831	4.47006	2.95962	6.12247	1.65831	C1
17	2.80617	5.48331	2.25391	6.76597	2.25391	C3
18	1.62801	3.53627	5.20072	6.02929	1.62801	C1
19	1.05534	4.61063	4.27602	6.95795	1.05534	C1
20	2.49499	6.35171	1.12060	8.05427	1.12060	C3
21	0.61498	4.58402	3.07802	6.63280	0.61498	C1
22	0.42012	4.71634	3.21984	6.86626	0.42012	C1
23	2.36957	3.79840	5.92184	6.39111	2.36957	C1
24	2.95406	7.25208	3.39755	9.51315	2.95406	C1
25	1.06731	3.97964	4.60404	6.39970	1.06731	C1

3.3.4 Lima Cluster

Nilai *centroid* dari perhitungan ROC yang diambil sebagai *centroid* awal untuk empat *cluster* adalah nilai yang memiliki nilai tertinggi, nilai terendah, dan nilai tengah. Nilai yang terpilih untuk *centroid* awal adalah data dengan nilai tertinggi di id 61, untuk data dengan nilai terendah di id 90, untuk data dengan nilai tengah kedua di id 3, untuk data dengan nilai tertinggi kedua di id 178, dan untuk data dengan nilai terendah di id 89 atau bisa dilihat pada Tabel 3.47.

Tabel 3.47 *Centroid* awal lima *cluster* dengan ROC.

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	11.42000	12.86000	0.86830	5.00800	2.85000	2.70000	4.00000
2	20.88000	17.05000	0.90310	6.45000	4.03200	5.01600	6.32100
3	14.29000	14.09000	0.90500	5.29100	3.33700	2.69900	4.82500
4	10.74000	12.73000	0.83290	5.14500	2.64200	4.70200	4.96300
5	21.18000	17.21000	0.89890	6.57300	4.03300	5.78000	6.23100

Tabel 3.48 Perhitungan di iterasi pertama untuk lima *cluster*.

Cluster	Hasil
1	4.60080
2	6.81793
3	1.45344
4	5.65121
5	7.45175

Langkah selanjutnya mencari *euclidean distance* dari setiap data *Seeds* dengan *centroid* awal yang telah ditentukan menggunakan persamaan (2.2) dan menggunakan *centroid* yang di Tabel 3.47.

Tabel 3.49 Hasil mengelompokkan lima *cluster* di iterasi pertama

NO	C1	C2	C3	C4	C5	TERDEKAT	CLUSTER
1	4.60080	6.81793	1.45344	5.65121	7.45175	1.45344	C3
2	4.37839	7.82876	1.86847	5.89442	8.51414	1.86847	C3
3	3.27857	7.85000	0.00000	4.35806	8.41450	0.00000	C3
4	2.87154	8.41590	0.64957	4.20077	8.99929	0.64957	C3
5	5.56198	6.50197	2.51926	6.82858	7.21698	2.51926	C3
6	3.45134	7.76867	0.32493	4.57914	8.35485	0.32493	C3
7	4.01201	7.03661	1.15875	4.53527	7.54823	1.15875	C3
8	3.18563	7.95225	0.28493	4.21400	8.51602	0.28493	C3
9	6.27161	5.48855	3.08082	7.18124	6.20254	3.08082	C3
10	5.91555	5.78421	2.71706	6.92304	6.48643	2.71706	C3
11	4.94783	6.23130	2.31239	5.07527	6.65969	2.31239	C3
12	3.28308	8.34360	1.05288	4.70969	8.96809	1.05288	C3
13	3.15482	7.95764	1.36751	3.54254	8.41126	1.36751	C3
14	2.87674	8.16670	0.72328	3.72139	8.68830	0.72328	C3
15	2.79774	8.26688	0.66748	3.77259	8.80128	0.66748	C3
16	3.90287	7.21080	1.52967	4.24850	7.65786	1.52967	C3
17	3.85692	7.90270	2.57248	3.55608	8.24973	2.57248	C3
18	4.98227	6.82992	1.93429	6.25540	7.51267	1.93429	C3
19	3.78062	7.84093	1.05178	5.22247	8.47293	1.05178	C3
20	2.25438	9.16388	2.19092	2.27198	9.59974	2.19092	C3
21	3.44978	7.66262	0.74529	4.20578	8.20237	0.74529	C3
22	3.32348	7.86533	0.54657	4.27064	8.43385	0.54657	C3
23	5.45820	7.08332	2.67035	6.90075	7.81432	2.67035	C3
24	1.77846	10.48222	2.73679	3.59717	11.07672	1.77846	C1
25	4.36948	7.27336	1.44248	5.62215	7.92042	1.44248	C3

Hasil perhitungan data pertama di iterasi pertama menggunakan *euclidean distance* dapat dilihat di Tabel 3.48. Hasil pengelompokkan menggunakan hasil perhitungan *euclidean distance* di iterasi pertama dapat dilihat di tabel 3.49 yang menampilkan contoh hasil akhir 25. Penentuan suatu data masuk ke *cluster* pertama atau kedua atau ketiga atau keempat atau kelima adalah dengan mencari hasil akhir perhitungan *euclidean distance* yang terdekat antar *cluster*. Sehingga dari iterasi pertama didapatkan jumlah anggota *cluster* pertama 17 data, *cluster* kedua 48 data, *cluster* ketiga 88 data, *cluster* keempat 54 data, dan *cluster* kelima 3 data.

Langkah selanjutnya mencari *centroid* baru dengan membagi jumlah semua data tiap *cluster* dengan jumlah anggota *cluster* di tiap atribut *dataset* atau menggunakan persamaan (2,1). Perhitungan *cluster* pertama dapat dilihat sebagai berikut :

$$V_{Area A} = \frac{203,31000}{17} = 11.95941$$

$$V_{Perimeter P} = \frac{224.09000}{17} = 13.18176$$

$$V_{Compactness C} = \frac{14.69950}{17} = 0.86468$$

$$V_{Length\ of\ Kernel} = \frac{87.27700}{17} = 5.13394$$

$$V_{Width\ of\ Kernel} = \frac{49.69100}{17} = 2.92300$$

$$V_{Asymmetry\ Coefficient} = \frac{47.37900}{17} = 2.78700$$

$$V_{Length\ of\ Kernel\ Groove} = \frac{82.37300}{17} = 4.84547$$

Perhitungan *centroid cluster* kedua, ketiga, keempat, dan kelima juga menggunakan persamaan (2,1) dan hasil akhir perhitungan pencarian *centroid* baru dapat dilihat di Tabel 3.50.

Tabel 3.50 Hasil *centroid* baru lima *cluster* di iterasi kedua

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	11.95941	13.18176	0.86468	5.13394	2.92300	2.78700	4.84547
2	18.97438	16.39917	0.88631	6.24710	3.75194	3.40410	6.10917
3	14.85807	14.56432	0.87866	5.60501	3.29327	2.98905	5.26284
4	11.77463	13.21870	0.84601	5.23085	2.83302	5.27031	5.11922
5	20.18667	16.91000	0.88693	6.38233	3.91067	6.12100	6.15733

Langkah selanjutnya mengelompokkan kembali dengan persamaan (2,2) menggunakan *centroid* baru. Hasil perhitungan data pertama di iterasi kedua menggunakan *centroid* baru dapat dilihat di Tabel 3.51.

Tabel 3.51 Hasil perhitungan di iterasi kedua untuk lima *cluster*

Cluster	Hasil
1	3.82771
2	4.34121
3	0.92447
4	4.95965
5	6.73704

Tabel 3.52 Hasil mengelompokkan lima *cluster* di iterasi kedua

NO	C1	C2	C3	C4	C5	TERDEKAT	CLUSTER
1	3.82771	4.34121	0.92447	4.95965	6.73704	0.92447	C3
2	3.73409	5.27152	1.99597	5.47118	7.88289	1.99597	C3
3	2.54212	5.52345	0.96162	3.74774	7.59794	0.96162	C3
4	2.15336	6.03429	1.50181	3.77651	8.20047	1.50181	C3
5	4.85703	3.93374	2.13963	6.18417	6.66244	2.13963	C3
6	2.69261	5.39010	0.88012	3.99247	7.56241	0.88012	C3
7	3.20043	4.85153	0.62848	3.64079	6.66877	0.62848	C3
8	2.39287	5.61791	0.98096	3.62065	7.69163	0.98096	C3
9	5.45576	2.90051	2.33454	6.37835	5.63851	2.33454	C3
10	5.13780	3.21371	2.05156	6.15156	5.90211	2.05156	C3
11	4.17351	4.32092	1.63558	3.97204	5.71061	1.63558	C3
12	2.56580	5.87310	1.60382	4.33500	8.22161	1.60382	C3
13	2.45942	5.88972	1.59353	2.66284	7.48126	1.59353	C3
14	2.09328	5.91616	1.27501	3.08423	7.81766	1.27501	C3
15	2.02571	5.98178	1.32546	3.20169	7.94540	1.32546	C3
16	3.20953	5.20251	1.37175	3.25764	6.72900	1.37175	C3
17	3.27874	6.16173	2.60971	2.39112	7.24166	2.39112	C4
18	4.28178	4.29781	1.66174	5.63151	6.88901	1.66174	C3
19	3.15429	5.38920	1.48725	4.73718	7.75955	1.48725	C3
20	1.57718	7.11154	2.66904	1.57207	8.63764	1.57207	C4
21	2.61424	5.37511	0.74761	3.49163	7.34961	0.74761	C3
22	2.47913	5.51482	0.87620	3.66238	7.60984	0.87620	C3
23	4.80086	4.50329	2.48571	6.37068	7.29092	2.48571	C3
24	1.38347	8.05203	3.52855	3.87431	10.26818	1.38347	C1
25	3.65321	4.76451	1.26554	5.04425	7.23201	1.26554	C3

Hasil akhir perhitungan *euclidean distance* di iterasi kedua dapat dilihat di tabel 3.52 yang menampilkan contoh hasil akhir 25 data pertama. Penentuan suatu data masuk ke *cluster* pertama atau kedua atau ketiga atau keempat atau kelima adalah dengan mencari hasil akhir perhitungan *euclidean distance* yang terdekat antar *cluster*. Sehingga dari iterasi kedua didapatkan jumlah anggota *cluster* pertama 29 data, *cluster* kedua 51 data, *cluster* ketiga 64 data, *cluster* keempat 57 data, dan *cluster* kelima 9 data. Dikarenakan anggota *cluster* sebelumnya dengan anggota *cluster* baru tidak sama atau berubah, maka proses *clustering* akan dilanjutkan dengan mengulangi kembali proses *clustering* dalam pencarian *centroid* baru menggunakan anggota *cluster* terakhir.

Langkah mencari *centroid* baru dengan menggunakan persamaan (2,1). Perhitungan *cluster* pertama dapat dilihat sebagai berikut :

$$V_{Area A} = \frac{354.15000}{29} = 12.21207$$

$$V_{Perimeter P} = \frac{386.62000}{29} = 13.33172$$

$$V_{Compactness C} = \frac{25.02430}{29} = 0.86291$$

$$V_{Length\ of\ Kernel} = \frac{150.90000}{29} = 5.20345$$

$$V_{Width\ of\ Kernel} = \frac{85.09600}{29} = 2.9343$$

$$V_{Asymmetry\ Coefficient} = \frac{80.78710}{29} = 2.78576$$

$$V_{Length\ of\ Kernel\ Groove} = \frac{142.40100}{29} = 4.91038$$

Perhitungan *centroid cluster* kedua, ketiga, keempat dan kelima juga menggunakan persamaan (2,1) dan hasil akhir perhitungan pencarian *centroid* baru dapat dilihat di Tabel 3.53.

Tabel 3.53 Hasil *centroid* baru lima *cluster* di iterasi ketiga

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	12.21207	13.33172	0.86291	5.20345	2.93434	2.78576	4.91038
2	18.52686	16.22373	0.88408	6.19222	3.69541	3.24082	6.05194
3	14.95984	14.60766	0.88048	5.61688	3.31567	2.80874	5.24669
4	11.94561	13.28316	0.84959	5.24109	2.86653	5.29718	5.11074
5	20.07000	16.82111	0.89109	6.34078	3.90556	5.44500	6.16700

Langkah selanjutnya mengelompokkan kembali dengan persamaan (2,2) menggunakan *centroid* baru. Hasil perhitungan data pertama di iterasi ketiga menggunakan *centroid* baru dapat dilihat di Tabel 3.54.

Tabel 3.54 Hasil perhitungan di iterasi ketiga untuk lima *cluster*

Cluster	Hasil
1	3.52638
2	3.82762
3	0.71532
4	4.83270
5	6.24807

Tabel 3.55 Hasil mengelompokkan lima *cluster* di iterasi ketiga

NO	C1	C2	C3	C4	C5	TERDEKAT	CLUSTER
1	3.52638	3.82762	0.71532	4.83270	6.24807	0.71532	C3
2	3.47281	4.76604	1.81749	5.37758	7.34952	1.81749	C3
3	2.25370	5.02432	1.00686	3.63403	7.18570	1.00686	C3
4	1.87671	5.52765	1.51232	3.68995	7.77370	1.51232	C3
5	4.57141	3.44211	1.92882	6.05900	6.09052	1.92882	C3
6	2.39900	4.88429	0.86753	3.88187	7.12901	0.86753	C3
7	2.90729	4.37340	0.83527	3.49092	6.30525	0.83527	C3
8	2.09538	5.11557	1.04484	3.51180	7.28252	1.04484	C3
9	5.15186	2.39301	2.17180	6.23670	5.06375	2.17180	C3
10	4.83729	2.70578	1.87026	6.01211	5.33844	1.87026	C3
11	3.89887	3.90310	1.78214	3.79674	5.42183	1.78214	C3
12	2.29516	5.35916	1.53701	4.25379	7.75449	1.53701	C3
13	2.21034	5.42701	1.78324	2.51652	7.17787	1.78324	C3
14	1.79949	5.42378	1.40854	2.97275	7.44990	1.40854	C3
15	1.72892	5.48475	1.43210	3.09895	7.56495	1.43210	C3
16	2.95134	4.75636	1.55771	3.09262	6.42680	1.55771	C3
17	3.10345	5.75775	2.81033	2.20854	7.04594	2.20854	C4
18	3.99710	3.79514	1.45092	5.50971	6.35122	1.45092	C3
19	2.89012	4.88555	1.36438	4.63530	7.27809	1.36438	C3
20	1.43560	6.64455	2.84620	1.47749	8.36318	1.43560	C1
21	2.30926	4.87833	0.89118	3.36887	6.96228	0.89118	C3
22	2.17337	5.00944	0.94324	3.55352	7.19713	0.94324	C3
23	4.53100	4.02196	2.27372	6.26006	6.70726	2.27372	C3
24	1.38574	7.53793	3.55360	3.89100	9.84450	1.38574	C1
25	3.36237	4.25242	1.07551	4.93159	6.73086	1.07551	C3

Hasil akhir perhitungan *euclidean distance* di iterasi ketiga dapat dilihat di Tabel 3.55 yang menampilkan contoh hasil akhir 25 data pertama. Penentuan suatu data masuk ke *cluster* pertama atau kedua atau ketiga atau keempat atau kelima adalah dengan mencari hasil akhir perhitungan *Euclidean distance* yang terdekat antar *cluster*. Sehingga dari iterasi ketiga didapatkan jumlah anggota *cluster* pertama 33 data, *cluster* kedua 50 data, *cluster* ketiga 60 data, *cluster* keempat 56 data, dan *cluster* kelima 11 data. Dikarenakan anggota *cluster* sebelumnya dengan anggota *cluster* baru tidak sama atau berubah, maka proses *clustering* akan dilanjutkan dengan mengulangi kembali proses *clustering* dalam pencarian *centroid* baru menggunakan anggota *cluster* terakhir.

Langkah mencari *centroid* baru dengan menggunakan persamaan (2,1). Perhitungan *cluster* pertama dapat dilihat sebagai berikut :

$$V_{Area A} = \frac{409.31000}{33} = 12.40333$$

$$V_{Perimeter P} = \frac{442.89000}{33} = 13.42091$$

$$V_{Compactness C} = \frac{28.53730}{33} = 0.86477$$

$$V_{Length\ of\ Kernel} = \frac{172.64500}{33} = 5.23167$$

$$V_{Width\ of\ Kernel} = \frac{97.79000}{33} = 2.96333$$

$$V_{Asymmetry\ Coefficient} = \frac{93.42810}{33} = 2.83115$$

$$V_{Length\ of\ Kernel\ Groove} = \frac{162.85300}{33} = 4.93494$$

Perhitungan *centroid cluster* kedua, ketiga, keempat dan kelima juga menggunakan persamaan (2,1) dan hasil akhir perhitungan pencarian *centroid* baru dapat dilihat di Tabel 3.56.

Tabel 3.56 Hasil *centroid* baru lima *cluster* di iterasi keempat

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	12.40333	13.42091	0.86477	5.23167	2.96333	2.83115	4.93494
2	18.46000	16.19520	0.88394	6.18240	3.68804	3.22884	6.04316
3	15.00683	14.63050	0.88060	5.62317	3.32262	2.76434	5.24095
4	11.89696	13.26054	0.84904	5.23593	2.85852	5.31545	5.11270
5	19.91182	16.76182	0.89030	6.32955	3.88009	5.30700	6.17036

Langkah selanjutnya mengelompokkan kembali dengan persamaan (2.2) menggunakan *centroid* baru. Hasil perhitungan data pertama di iterasi keempat menggunakan *centroid* baru dapat dilihat di Tabel 3.57.

Tabel 3.57 Hasil perhitungan di iterasi keempat untuk lima *cluster*

Cluster	Hasil
1	3.32142
2	3.75331
3	0.65069
4	4.88628
5	6.03353

Tabel 3.58 Hasil mengelompokkan lima *cluster* di iterasi keempat

NO	C1	C2	C3	C4	C5	TERDEKAT	CLUSTER
1	3.32142	3.75331	0.65069	4.88628	6.03353	0.65069	C3
2	3.31407	4.69557	1.77638	5.42523	7.13403	1.77638	C3
3	2.04487	4.95012	1.04614	3.68487	6.98061	1.04614	C3
4	1.69114	5.45311	1.54163	3.73556	7.56613	1.54163	C3
5	4.38213	3.37508	1.86089	6.11223	5.87600	1.86089	C3
6	2.19383	4.80969	0.89382	3.93242	6.92059	0.89382	C3
7	2.68695	4.30060	0.89633	3.54728	6.10483	0.89633	C3
8	1.88313	5.04089	1.09076	3.56200	7.07630	1.09076	C3
9	4.94594	2.32313	2.11117	6.29190	4.84494	2.11117	C3
10	4.63257	2.63511	1.80470	6.06729	5.12121	1.80470	C3
11	3.68609	3.83755	1.81552	3.85352	5.23431	1.81552	C3
12	2.13009	5.28489	1.54245	4.29875	7.54168	1.54245	C3
13	2.01013	5.35583	1.84789	2.57206	6.98810	1.84789	C3
14	1.58081	5.34973	1.46745	3.02344	7.24886	1.46745	C3
15	1.51332	5.41041	1.48542	3.14803	7.36216	1.48542	C3
16	2.74376	4.68718	1.61275	3.14970	6.23918	1.61275	C3
17	2.94515	5.69331	2.87138	2.26170	6.87840	2.26170	C4
18	3.80675	3.72458	1.38691	5.56234	6.13702	1.38691	C3
19	2.71431	4.81299	1.34636	4.68395	7.06807	1.34636	C3
20	1.32113	6.57267	2.91260	1.52376	8.17615	1.32113	C1
21	2.09042	4.80390	0.95306	3.42172	6.75734	0.95306	C3
22	1.96037	4.93443	0.99122	3.60392	6.98908	0.99122	C3
23	4.36049	3.95754	2.21107	6.31091	6.49318	2.21107	C3
24	1.47157	7.46308	3.58486	3.91101	9.63470	1.47157	C1
25	3.17240	4.17937	1.02737	4.98288	6.51620	1.02737	C3

Hasil akhir perhitungan *euclidean distance* di iterasi ketiga dapat dilihat di Tabel 3.58 yang menampilkan contoh hasil akhir 25 data pertama. Penentuan suatu data masuk ke *cluster* pertama atau kedua atau ketiga atau keempat atau kelima adalah dengan mencari hasil akhir perhitungan *euclidean distance* yang terdekat antar *cluster*. Sehingga dari iterasi keempat didapatkan jumlah anggota *cluster* pertama 33 data, *cluster* kedua 49 data, *cluster* ketiga 60 data, *cluster* keempat 56 data, dan *cluster* kelima 12 data. Dikarenakan anggota *cluster* sebelumnya dengan anggota *cluster* baru tidak sama atau berubah, maka proses *clustering* akan dilanjutkan dengan mengulangi kembali proses *clustering* dalam pencarian *centroid* baru menggunakan anggota *cluster* terakhir.

Langkah mencari *centroid* baru dengan menggunakan persamaan (2,1). Perhitungan *cluster* pertama dapat dilihat sebagai berikut :

$$V_{\text{Area A}} = \frac{409.31000}{33} = 12.40333$$

$$V_{\text{Perimeter P}} = \frac{442.89000}{33} = 13.42091$$

$$V_{\text{Compactness C}} = \frac{28.53730}{33} = 0.86477$$

$$V_{\text{Length of Kernel}} = \frac{172.64500}{33} = 5.23167$$

$$V_{\text{Width of Kernel}} = \frac{97.79000}{33} = 2.96333$$

$$V_{\text{Asymmetry Coefficient}} = \frac{93.42810}{33} = 2.83115$$

$$V_{\text{Length of Kernel Groove}} = \frac{162.85300}{33} = 4.93494$$

Perhitungan *centroid cluster* kedua, ketiga, keempat dan kelima juga menggunakan persamaan (2,1) dan hasil akhir perhitungan pencarian *centroid* baru dapat dilihat di Tabel 3.59.

Tabel 3.59 Hasil *centroid* baru lima *cluster* di iterasi kelima

Centroid	Area A	Perimeter P	Compactness C	Length of kernel	Width of Kernel	Asymmetry Coefficient	Length of Kernel Groove
1	12.40333	13.42091	0.86477	5.23167	2.96333	2.83115	4.93494
2	18.46204	16.18857	0.88473	6.17253	3.69218	3.19406	6.03490
3	15.00683	14.63050	0.88060	5.62317	3.32262	2.76434	5.24095
4	11.89696	13.26054	0.84904	5.23593	2.85852	5.31545	5.11270
5	19.78250	16.74167	0.88654	6.35758	3.84717	5.27583	6.19350

Tabel 3.60 Hasil perhitungan di iterasi kelima untuk lima *cluster*

Cluster	Hasil
1	3.32142
2	3.74100
3	0.65069
4	4.88628
5	5.91517

Tabel 3.61 Hasil mengelompokkan lima *cluster* di iterasi kelima

NO	C1	C2	C3	C4	C5	TERDEKAT	CLUSTER
1	3.32142	3.74100	0.65069	4.88628	5.91517	0.65069	C3
2	3.31407	4.67565	1.77638	5.42523	7.02292	1.77638	C3
3	2.04487	4.94192	1.04614	3.68487	6.86367	1.04614	C3
4	1.69114	5.44284	1.54163	3.73556	7.44845	1.54163	C3
5	4.38213	3.35143	1.86089	6.11223	5.77213	1.86089	C3
6	2.19383	4.80011	0.89382	3.93242	6.80264	0.89382	C3
7	2.68695	4.90023	0.89633	3.54728	5.98215	0.89633	C3
8	1.88313	5.03350	1.09076	3.56200	6.95639	1.09076	C3
9	4.94594	2.30434	2.11117	6.29190	4.73188	2.11117	C3
10	4.63257	2.61544	1.80470	6.06729	5.01100	1.80470	C3
11	3.68609	3.84673	1.81552	3.85352	5.11104	1.81552	C3
12	2.13009	5.27158	1.54245	4.29875	7.42340	1.54245	C3
13	2.01013	5.35693	1.84789	2.57206	6.86841	1.84789	C3
14	1.58081	5.34571	1.46745	3.02344	7.12715	1.46745	C3
15	1.51332	5.40508	1.48542	3.14803	7.24069	1.48542	C3
16	2.74376	4.68974	1.61275	3.14970	6.12247	1.61275	C3
17	2.94515	5.70108	2.87138	2.26170	6.76597	2.26170	C4
18	3.80675	3.70469	1.38691	5.56234	6.02929	1.38691	C3
19	2.71431	4.79720	1.34636	4.68395	6.95795	1.34636	C3
20	1.32113	6.57408	2.91260	1.52376	8.05427	1.32113	C1
21	2.09042	4.80018	0.95306	3.42172	6.63280	0.95306	C3
22	1.96037	4.92769	0.99122	3.60392	6.86626	0.99122	C3
23	4.36049	3.93194	2.21107	6.31091	6.39111	2.21107	C3
24	1.47157	7.45161	3.58486	3.91101	9.51315	1.47157	C1
25	3.17240	4.16442	1.02737	4.98288	6.39970	1.02737	C3

Langkah selanjutnya mengelompokkan kembali dengan persamaan (2,2) menggunakan *centroid* baru. Sebagai contoh perhitungan data pertama di iterasi ketujuh menggunakan *centroid* baru dapat dilihat di Tabel 3.60. Hasil akhir perhitungan *euclidean distance* di iterasi ketiga dapat dilihat di Tabel 3.10 yang menampilkan contoh hasil akhir 25 data pertama.

Penentuan suatu data masuk ke *cluster* pertama atau kedua atau ketiga atau keempat atau kelima adalah dengan mencari hasil akhir perhitungan *euclidean distance* yang terdekat antar *cluster*. Sehingga dari iterasi keempat didapatkan jumlah anggota *cluster* pertama 33 data, *cluster* kedua 49 data, dan *cluster* ketiga 60 data, *cluster* keempat 56 data, dan *cluster* kelima 12 data. Proses *clustering* berhenti di iterasi kelima karena anggota *cluster* sebelumnya dengan anggota *cluster* baru sama atau tidak mengalami perubahan. Sehingga proses *clustering* lima *cluster* dengan penentuan *centroid* awal menggunakan penentuan metode ROC menghasilkan *cluster* pertama dengan anggota 33 data, *cluster* kedua dengan anggota 49 data, *cluster* ketiga dengan anggota 60 data, *cluster* ketiga dengan anggota 56 data, dan *cluster* keempat dengan anggota 12 data.

3.4 Perhitungan Nilai DBI di K-Means dan ROC

3.4.1 Dua Cluster

Proses setelah *clustering* adalah menghitung nilai validasi menggunakan metode DBI. Langkah awal perhitungan menggunakan metode DBI dengan menghitung nilai SSW.

Tabel 3.62 Hasil nilai jarak di 2 *cluster*.

<i>Cluster</i>	Nilai Jarak (d)
1	280.55009
2	141.84488

Mencari nilai SSW, pertama menghitung jarak (d) setiap data *cluster* ke *centroid* menggunakan rumus *euclidean distance* kemudian menjumlahkan semua nilai jarak (d) dan selanjutnya dibagi dengan jumlah

data dalam *cluster* atau menggunakan persamaan (2,4). Setelah nilai jarak (d) untuk *cluster* pertama dan kedua didapatkan langkah selanjutnya menghitung jumlah semua nilai jarak (d), dan diperoleh untuk *cluster* pertama 280.55009 sedangkan untuk *cluster* kedua 141.84488 atau bisa dilihat di Tabel 3.62. Langkah selanjutnya menghitung nilai SSW menggunakan persamaan (2,4) dan perhitungan SSW sebagai berikut :

$$SSW_1 = \frac{280.55009}{134} = 2.09366$$

$$SSW_2 = \frac{141.84488}{76} = 1.86638$$

Langkah selanjutnya mencari nilai SSB dengan cara menghitung jarak *centroid* ke *centroid* menggunakan rumus *euclidean distance* atau menggunakan persamaan (2,5) dan hasil SSB dapat dilihat di Tabel 3.63.

Tabel 3.63 Hasil perhitungan SSB dua *cluster* di K-Means dan ROC

Centroid	Hasil
1 & 2	5.93714

Langkah berikutnya mencari nilai Rasio dengan menjumlahkan SSW *cluster* pertama dengan SSW *cluster* kedua kemudian dibagi dengan nilai SSB atau menggunakan persamaan (2,6). Perhitungan Rasio sebagai berikut :

$$\text{Rasio} = \frac{(2.09366 + 1.86638)}{5.93714} = 0.66699$$

Langkah terakhir menghitung nilai DBI dengan menjumlahkan semua nilai maksimal rasio dan dibagi dengan jumlah *cluster* yang digunakan atau menggunakan persamaan (2,7). Perhitungan DBI sebagai berikut :

$$\text{DBI} = \frac{0.66699}{2} = 0.33350$$

3.4.2 Tiga Cluster

Proses setelah *clustering* adalah menghitung nilai validasi menggunakan metode DBI. Langkah awal perhitungan menggunakan metode DBI dengan menghitung nilai SSW. Untuk mencari nilai SSW, pertama menghitung jarak (d) setiap data *cluster* ke *centroid* menggunakan rumus *euclidean distance* kemudian menjumlahkan semua nilai jarak (d) dan selanjutnya dibagi dengan jumlah data dalam *cluster* atau menggunakan persamaan (2,4).

Tabel 3.64 Hasil nilai jarak di 3 *cluster*.

<i>Cluster</i>	Nilai Jarak (d)
1	122.49732
2	95.05401
3	96.73806

Setelah nilai jarak (d) untuk *cluster* pertama, kedua, dan ketiga didapatkan langkah selanjutnya menghitung jumlah semua nilai jarak (d), dan diperoleh untuk *cluster* pertama 122.49732, untuk *cluster* kedua 95.05401, dan untuk *cluster* ketiga 96.73806.

Langkah selanjutnya menghitung nilai SSW menggunakan persamaan (2,4) dan perhitungan SSW sebagai berikut :

$$SSW_1 = \frac{122.49732}{82} = 1.49387$$

$$SSW_2 = \frac{95.05401}{61} = 1.55826$$

$$SSW_3 = \frac{96.73806}{67} = 1.44385$$

Langkah selanjutnya mencari nilai SSB dengan cara menghitung jarak *centroid* ke *centroid* menggunakan rumus *euclidean distance* atau menggunakan persamaan (2,5) dan hasil SSB dapat dilihat di Tabel 3.65.

Tabel 3.65 Hasil perhitungan SSB tiga *cluster* di K-Means dan ROC

Centroid	Hasil
1 & 2	7.62048
1 & 3	3.66792
2 & 3	4.52035

Langkah berikutnya mencari nilai Rasio dengan menjumlahkan SSW *cluster* pertama dengan SSW *cluster* kedua kemudian dibagi dengan nilai SSB atau menggunakan persamaan (2,6). Perhitungan Rasio sebagai berikut :

Cluster 1 dan 2

$$\frac{(1.49387 + 1.55826)}{7.62048} = 0.40052$$

Cluster 1 dan 3

$$\frac{(1.49387 + 1.44385)}{3.66792} = 0.80092$$

Cluster 2 dan 3

$$\frac{(1.55826 + 1.44385)}{4.52035} = 0.66413$$

Langkah terakhir menghitung nilai DBI dengan menjumlahkan semua nilai maksimal Rasio tiap *cluster* dibagi dengan jumlah *cluster* yang digunakan atau menggunakan persamaan (2,7). Nilai Rasio tertinggi untuk tiap *cluster* adalah *cluster* 1 dan 3 dengan nilai 0.80092, *cluster* 2 dan 3 dengan nilai 0.66413, dan *cluster* 3 dan 1 dengan nilai 0.80092. Perhitungan DBI sebagai berikut :

$$DBI = \frac{2.26598}{3} = 0.75533$$

3.4.3 Empat *Cluster*

Proses setelah *clustering* adalah menghitung nilai validasi menggunakan metode DBI. Langkah awal perhitungan menggunakan metode

DBI dengan menghitung nilai SSW. Untuk mencari nilai SSW, pertama menghitung jarak (d) setiap data *cluster* ke *centroid* menggunakan rumus *euclidean distance* kemudian menjumlahkan semua nilai jarak (d) dan selanjutnya dibagi dengan jumlah data dalam *cluster* atau menggunakan persamaan (2,4).

Tabel 3.66 Hasil nilai jarak di 4 *cluster*.

<i>Cluster</i>	Nilai Jarak (d)
1	101.50415
2	76.96771
3	104.24577
4	14.95989

Setelah nilai jarak (d) untuk *cluster* pertama, kedua, ketiga dan keempat didapatkan langkah selanjutnya menghitung jumlah semua nilai jarak (d), dan diperoleh untuk *cluster* pertama 101.50415, untuk *cluster* kedua 76.96771 untuk *cluster* ketiga 104.24577, untuk *cluster* keempat 14.95989. Langkah selanjutnya menghitung nilai SSW menggunakan persamaan (2,4) dan perhitungan SSW sebagai berikut :

$$SSW_1 = \frac{101.50415}{69} = 1.47107$$

$$SSW_2 = \frac{76.96771}{54} = 1.42533$$

$$SSW_3 = \frac{104.24577}{75} = 1.38994$$

$$SSW_4 = \frac{14.95989}{12} = 1.24666$$

Langkah selanjutnya mencari nilai SSB dengan cara menghitung jarak *centroid* ke *centroid* menggunakan rumus *euclidean distance* atau menggunakan persamaan (2,5) hasil SSB dapat dilihat di Tabel 3.67.

Tabel 3.67 Hasil perhitungan SSB empat *cluster* di K-Means dan ROC

Centroid	Hasil
1 & 2	4.37323
1 & 3	3.58444
1 & 4	6.57300
2 & 3	7.26937
2 & 4	2.63703
3 & 4	8.79041

Langkah berikutnya mencari nilai Rasio dengan menjumlahkan SSW *cluster* pertama dengan SSW *cluster* kedua kemudian dibagi dengan nilai SSB atau menggunakan persamaan (2,6). Perhitungan Rasio sebagai berikut :

Cluster 1 dan 2

$$\frac{(1.47107 + 1.42533)}{4.37323} = 0.66230$$

Cluster 1 dan 3

$$\frac{(1.47107 + 1.38994)}{3.58444} = 0.79818$$

Cluster 1 dan 4

$$\frac{(1.47107 + 1.24666)}{6.57300} = 0.41347$$

Cluster 2 dan 3

$$\frac{(1.42533 + 1.38994)}{7.26937} = 7.26937$$

Cluster 2 dan 4

$$\frac{(1.42533 + 1.24666)}{2.63703} = 2.63703$$

Cluster 3 dan 4

$$\frac{(1.38994 + 1.24666)}{8.79041} = 8.79041$$

Langkah terakhir menghitung nilai DBI dengan menjumlahkan semua nilai maksimal Rasio tiap *cluster* dibagi dengan jumlah *cluster* yang digunakan atau menggunakan persamaan (2,7). Nilai Rasio tertinggi untuk tiap *cluster* adalah *cluster* 1 dan 3 dengan nilai 0.79818, *cluster* 2 dan 4 dengan nilai 1.01325, *cluster* 3 dan 1 dengan nilai 0.79818, dan *cluster* 4 dan 2 dengan nilai 1.01325. Perhitungan DBI sebagai berikut :

$$DBI = \frac{3.62287}{4} = 0.72457$$

3.4.4 Lima Cluster

Proses setelah *clustering* adalah menghitung nilai validasi menggunakan metode DBI. Langkah awal perhitungan menggunakan metode DBI dengan menghitung nilai SSW. Untuk mencari nilai SSW, pertama menghitung jarak (d) setiap data *cluster* ke *centroid* menggunakan rumus *euclidean distance* kemudian menjumlahkan semua nilai jarak (d) dan selanjutnya dibagi dengan jumlah data dalam *cluster* atau menggunakan persamaan (2,4).

Tabel 3.68 Hasil nilai jarak di 5 *cluster*

<i>Cluster</i>	Nilai Jarak (d)
1	36.56551
2	63.24515
3	82.52175
4	70.01626
5	14.15769

Setelah nilai jarak (d) untuk *cluster* pertama, kedua, ketiga, keempat, dan kelima didapatkan langkah selanjutnya menghitung jumlah semua nilai jarak (d), dan diperoleh untuk *cluster* pertama 36.56551, untuk *cluster* kedua 63.24515 untuk *cluster* ketiga 82.52175 untuk *cluster* keempat 70.01626, dan untuk *cluster* kelima 14.15769.

Langkah selanjutnya menghitung nilai SSW menggunakan persamaan (2,4) dan perhitungan SSW sebagai berikut :

$$SSW_1 = \frac{36.56551}{33} = 1.13807$$

$$SSW_2 = \frac{63.24515}{49} = 1.29072$$

$$SSW_3 = \frac{82.52175}{49} = 1.37536$$

$$SSW_3 = \frac{70.01626}{56} = 1.25029$$

$$SSW_3 = \frac{14.15769}{12} = 1.24666$$

Langkah selanjutnya mencari nilai SSB dengan cara menghitung jarak *centroid* ke *centroid* menggunakan rumus *euclidean distance* atau menggunakan persamaan (2,5) dan hasil SSB dapat dilihat di Tabel 3.69.

Tabel 3.69 Hasil perhitungan SSB tiga *cluster* di K-Means dan ROC

Centroid	Hasil
1 & 2	6.86486
1 & 3	2.93683
1 & 4	2.54886
1 & 5	8.66541
2 & 3	3.95215
2 & 4	7.65492
2 & 5	2.54298
3 & 4	4.29407
3 & 5	5.94084
4 & 5	8.81513

Langkah berikutnya mencari nilai Rasio dengan menjumlahkan SSW *cluster* pertama dengan SSW *cluster* kedua kemudian dibagi dengan

nilai SSB atau menggunakan persamaan (2,6). Perhitungan Rasio sebagai berikut :

Cluster 1 dan 2

$$\frac{(1.13807 + 1.29072)}{6.86486} = 0.35380$$

Cluster 1 dan 3

$$\frac{(1.13807 + 1.37536)}{2.93633} = 0.85598$$

cluster 1 dan 4

$$\frac{(1.13807 + 1.25029)}{2.54886} = 0.93703$$

cluster 1 dan 5

$$\frac{(1.13807 + 1.24666)}{8.66541} = 0.27520$$

Cluster 2 dan 3

$$\frac{(1.29072 + 1.37536)}{3.95215} = 0.67459$$

Cluster 2 dan 4

$$\frac{(1.29072 + 1.25029)}{7.65492} = 0.33194$$

Cluster 2 dan 5

$$\frac{(1.29072 + 1.24666)}{2.54298} = 0.99780$$

Cluster 3 dan 4

$$\frac{(1.37536 + 1.25029)}{4.29407} = 0.61146$$

Cluster 3 dan 5

$$\frac{(1.37536 + 1.24666)}{5.94084} = 0.44136$$

Cluster 4 dan 5

$$\frac{(1.25029 + 1.24666)}{8.81513} = 0.28326$$

Langkah terakhir menghitung nilai DBI dengan menjumlahkan semua nilai maksimal Rasio tiap *cluster* dibagi dengan jumlah *cluster* yang

digunakan atau menggunakan persamaan (2,7). Nilai Rasio tertinggi untuk tiap *cluster* adalah *cluster* 1 dan 4 dengan nilai 0.93703, *cluster* 2 dan 5 dengan nilai 0.99780, *cluster* 3 dan 1 dengan nilai 0.85598, *cluster* 4 dan 1 dengan nilai 0.93703, dan *cluster* 5 dan 2 dengan nilai 0.99780. Perhitungan DBI sebagai berikut :

$$DBI = \frac{4.72563}{5} = 0.94513$$

