

BAB 4

HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

4.1.1 Deskripsi Data

Sebelum proses *clustering*, dilakukan pengumpulan semua data dari Badan Pusat Statistik (BPS) Sumatera Utara mengenai indikator kemiskinan berdasarkan kabupaten/kota di Sumatera Utara tahun 2023. Berikut merupakan data yang digunakan:

Tabel 4.1 : Data Penelitian

Kabupaten/Kota	TPT	PPM	P_1	P_2
Nias	2,31	15,1	1,95	0,4
Mandailing Natal	7,45	8,86	1,48	0,44
Tapanuli Selatan	3,49	7,01	0,92	0,18
Tapanuli Tengah	7,81	11,5	1,58	0,3
Tapanuli Utara	1,03	8,54	0,94	0,18
Toba	1,3	8,04	1,02	0,18
Labuhanbatu	5,99	7,99	1,4	0,37
Asahan	6,12	8,21	0,77	0,13
Simalungun	5,35	7,87	0,99	0,21
Dairi	1,23	7,47	0,93	0,17
Karo	2,63	7,98	1,17	0,29
Deli Serdang	8,62	3,44	0,34	0,07
Langkat	6,33	9,23	2,22	0,79
Nias Selatan	3,48	16,39	3,04	0,84
Humbang Hasundutan	0,84	8,69	0,99	0,19
Pakpak Bharat	0,45	7,54	0,49	0,07
Samosir	1,03	11,66	1,87	0,49
Serdang Bedagai	4,97	7,44	0,89	0,21
Batu Bara	5,88	11,38	1,51	0,36
Padang Lawas Utara	4,42	8,79	1,14	0,24
Padang Lawas	5,75	7,89	1,44	0,42
Labuhanbatu Selatan	3,43	8,06	0,94	0,18
Labuanbatu Utara	4,84	9,08	1,19	0,27
Nias Utara	2,57	21,79	2,52	0,59
Nias Barat	0,8	22,81	2,08	0,34
Sibolga	6,79	11,42	1,49	0,29
Tanjungbalai	4,47	12,21	1,34	0,24
Pematangsiantar	8,62	7,24	1	0,2
Tebing Tinggi	6,24	9,49	1,49	0,35
Medan	8,67	8	0,92	0,19
Binjai	6,1	4,79	0,48	0,06
Padangsidempuan	7,57	6,85	0,87	0,17
Gunungsitoli	3,67	14,78	2,08	0,46

Keterangan:

TPT : Tingkat Pengangguran Terbuka (X_1)

PPM : Persentase Penduduk Miskin (X_2)

P_1 : Indeks Kedalaman Kemiskinan (X_3)

P_2 : Indeks Keparahan Kemiskinan (X_4)

Berikut adalah deskripsi statistik yang digunakan untuk melihat gambaran dari data.

Tabel 4.2 : Statistika Deskriptif

Statistika	X_1	X_3	X_4	X_5
Minimum	0,45	3,44	0,34	0,13
Maksimum	8,67	754	3,04	0,84
Mean	4,5530	32,5455	1,3176	0,3536
Std. Deviasi	2,55136	129,58089	0,60101	0,19591

4.1.2 Uji Multikolinearitas

Sebelum dilakukan pengklasteran, dilakukan uji asumsi Multikolinearitas untuk mengetahui ukuran kemiripan apa yang dapat digunakan. Cara untuk mengetahui adanya multikolinearitas yaitu dengan melihat nilai Tolerance dan VIF (*Variance Inflation Factor*). Pedoman keputusan untuk nilai *Tolerance* ialah jika nilai *Tolerance* $> 0,10$ maka artinya tidak terjadi multikolinearitas. Jika nilai *Tolerance* $< 0,10$ maka artinya terjadi multikolinearitas. Kemudian pedoman yang digunakan dalam melihat nilai VIF ialah jika nilai VIF < 10 maka tidak terjadi multikolinearitas sedangkan jika nilai VIF > 10 maka artinya terjadi multikolinearitas. Berikut hasil pengujian multikolinearitas dengan X_2 sebagai variabel *dependent*:

Tabel 4.3 : Nilai VIF

Variabel	Tolerance	VIF
X_1	0,623	1,604
X_2	0,567	1,764
X_3	0,543	1,840
X_4	0,536	1,865

Berdasarkan Tabel 4.3 dapat dilihat bahwasanya nilai *Tolerance* $> 0,10$ dan nilai VIF dari setiap variabel < 10 . Sehingga dapat disimpulkan bahwa tidak terjadi multikolinearitas pada variabel-variabel tersebut. Oleh karena itu dalam

melakukan *clustering* pada metode *k-means* dapat menggunakan jarak *euclidean*.

4.1.3 Uji Signifikansi

Tabel 4.4 : Nilai Sig.

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	t
		B	Std. Error	Beta	
1	(Constant)	146.846	60.537		2.426
	X1	-20.336	7.260	-.400	-2.801
	X3	-118.840	34.534	-.551	-3.441
	X4	381.384	104.327	.577	3.656
a. Dependent Variable: X ₂					

Pada penelitian ini pengujian dilakukan melalui pengamatan nilai signifikansi *t* pada tingkat *a* sebesar 5% dimana jika nilai signifikansi < 0,05, maka hipotesis teruji yang berarti independen berpengaruh signifikan terhadap variabel dependen. Pada tabel 4.4 dapat dilihat bahwa nilai sig. pada setiap variabel < 0,05 artinya hipotesis teruji,

4.1.4 Proses *Clustering* dengan *Agglomerative Hierarchical Clustering*

Langkah pertama yang dilakukan pada proses *clustering* ialah mencari jarak antar data dengan menghitung nilai *euclidean distance*, sebagai berikut dengan rumus pada persamaan 2.1.

Perhitungan d (Nias, y)

$$d(\text{Nias}, \text{Nias}) = \sqrt{(2,31 - 2,31)^2 + (15,1 - 15,1)^2 + (1,95 - 1,95)^2 + (0,4 - 0,4)^2} \\ = 0$$

$$d(\text{Nias}, \text{Mandailing Natal}) = \sqrt{(2,31 - 7,54)^2 + (15,1 - 8,86)^2 + (1,95 - 1,48)^2 + (0,4 - 0,44)^2} \\ = 8,10$$

$$d(\text{Nias}, \text{Tapanuli Selatan}) = \sqrt{(2,31 - 3,49)^2 + (15,1 - 7,01)^2 + (1,95 - 0,92)^2 + (0,4 - 0,18)^2} \\ = 8,24$$

Perhitungan dilakukan hingga kabupaten terakhir yaitu gunung sitoli, y (dapat dilihat pada lampiran 1). Jika hasil perhitungan jarak di atas dimasukkan dalam sebuah tabel, maka akan membentuk suatu matriks sebagai berikut:

Tabel 4.5 : Jarak *Euclidean K-Means Clustering*

Kab/Kota	Nias	Mandailing Natal	Tapanuli Selatan	Tapanuli Tengah	...	...	Gunung Sitoli
Nias	0	8,10	8,24	6,58	...	...	1,40
Mandailing Natal	8,10	0	4,41	2,67	...	...	7,05
Tapanuli Selatan	8,24	4,41	0	6,27	...	...	7,86
Tapanuli Tengah	6,58	2,67	6,27	0	...	...	5,31
...	...	...	...	...	...	...	...
Gunung Sitoli	1,40	7,05	7,86	5,31	...	...	0

Berdasarkan perhitungan *euclidean distance* yang telah dilakukan, maka dapat diketahui bahwa nilai terkecil terletak pada kabupaten/kota Tapanuli Utara dan Humbang Hasundutan dengan nilai 0,25. Sehingga langkah berikutnya ialah mengelompokkan kabupaten/kota menggunakan *mode single linkage* dengan rumus pada persamaan 2.3.

$$\begin{aligned} d(\text{Tapanuli Utara, Humbang Hasundutan}) \text{ Nias} &= \min\{6,76; 6,65\} \\ &= 6,65 \end{aligned}$$

$$\begin{aligned} d(\text{Tapanuli Utara, Humbang Hasundutan}) \text{ Mandailing Natal} &= \min\{6,46; 6,64\} \\ &= 6,46 \end{aligned}$$

$$\begin{aligned} d(\text{Tapanuli Utara, Humbang Hasundutan}) \text{ Tapanuli Selatan} &= \min\{2,90; 3,14\} \\ &= 2,90 \end{aligned}$$

Setelah dilakukan perhitungan hingga kabupaten terakhir menggunakan *mode single linkage* (lampiran 2) maka jarak terkecil pada *cluster* Tapanuli Utara dan Humbang Hasundutan menjadi 0,57. Berdasarkan perhitungan sebelumnya yang telah dilakukan jumlah *cluster* yang terbentuk ialah 32 *cluster* sehingga perhitungan menggunakan *mode single linkage* akan terus berlanjut hingga kabupaten/kota dikelompokkan menjadi 3 *cluster*. Selanjutnya nilai terkecil berada pada Labuhanbatu dan Padang Lawas dengan jarak 0,27.

$$d(\text{Labuhanbatu, Padang Lawas}) \text{ Nias} = \min\{8, 02; 8\}$$

$$= 8$$

$$d(\text{Labuhanbatu, Padang Lawas}) \text{ Mandailing Natal} = \min\{1, 70; 1, 96\}$$

$$= 1,70$$

$$d(\text{Labuhanbatu, Padang Lawas}) \text{ Tapanuli Selatan} = \min\{2, 73; 2, 49\}$$

$$= 2,49$$

Setelah dilakukan perhitungan terbentuk *cluster* kedua yaitu Labuhanbatu dan Padang Lawas dengan jarak terkecil berubah menjadi 0,64. Perhitungan dilakukan seperti langkah sebelumnya hingga *cluster* yang terbentuk menjadi 3 *cluster*. Hasil akhir dari pengelompokan menggunakan metode AHC didapat seperti pada tabel berikut:

Tabel 4.6 : Hasil *Clustering* Metode AHC

Kab/Kota	K_1, K_{33}	$K_2, K_3, K_{11}, K_{22}, K_{20}, K_{23}, K_5, K_{15}, K_6, K_9, K_{10}, K_{18}, K_7, K_{21}, K_8, K_{13}, K_{16}, K_{29}, K_{28}, K_{30}, K_{32}, K_4, K_{19}, K_{26}, K_{27}, K_{31}, K_{12}, K_{17}$	K_{14}, K_{24}, K_{25}
K_1, K_{33}	0	3,66	2,10
$K_2, K_3, K_{11}, K_{22}, K_{20}, K_{23}, K_5, K_{15}, K_6, K_9, K_{10}, K_{18}, K_7, K_{21}, K_8, K_{13}, K_{16}, K_{29}, K_{28}, K_{30}, K_{32}, K_4, K_{19}, K_{26}, K_{27}, K_{31}, K_{12}, K_{17}$	3,66	0	4,66
K_{14}, K_{24}, K_{25}	2,10	4,66	0

Keterangan : K_n adalah kabupaten/kota Pada pengelompokan menggunakan metode AHC, terbentuk 3 *cluster* dengan kabupaten/kota sebagai berikut:

Tabel 4.7 : Pengelompokkan Kabupaten/Kota berdasarkan Metode AHC

<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>
Nias, dan Gunung Sitoli	Mandailing Natal, Tapanuli Selatan, Karo, Labuhanbatu Selatan, Padang Lawas Utara, Labuhanbatu Utara, Tapanuli Utara, Humbang Hasundutan, Toba, Simalungun, Dairi, Serdang Bedagang, Labuhanbatu, Padang Lawas, Asahan, Langkat, Pakpak Barat, Tebing Tinggi, Pematang Siantar, Medan, Padang Sidimpuan, Tapanuli Tengah, Batu Bara, Sibolga, Tanjung Balai, Binjai, deli Serdang, dan Samosir.	Nias Selatan, Nias Utara, dan Nias Barat.

4.1.5 Proses *Clustering* dengan *K-Means*

Pengelompokkan penduduk miskin pada provinsi Sumatera Utara menggunakan algoritma *k-means* dimulai dengan menentukan jumlah *cluster* yang akan dibentuk lalu menentukan *centroid* awal dan seterusnya sebagai berikut:

1. Jumlah *cluster* yang akan dibentuk sebanyak 3 *cluster*.
2. Menentukan nilai pusat (*centroid*) dengan cara acak, berikut *centroid* awal yang dipilih:

Tabel 4.8 : *Centroid* Awal

<i>Centroid</i>	TPT	PPM	P_1	P_2
C_1	1,3	8,04	1,02	0,18
C_2	3,49	7,01	0,92	0,18
C_3	7,81	11,5	1,58	0,3

3. Menghitung jarak terdekat dengan *centroid* menggunakan jarak *euclidean*.

$$\begin{aligned}
 d(x_1, c_1) &= \sqrt{((2,31 - 1,3)^2 + (15,1 - 8,04)^2 + (1,95 - 1,02)^2 + (0,4 - 0,18)^2)} \\
 &= 7,20
 \end{aligned}$$

$$d(x_2, c_1) = \sqrt{((7,45 - 1,3)^2 + (8,86 - 8,04)^2 + (1,48 - 1,02)^2 + (0,44 - 0,18)^2)}$$

$$= 6,23$$

$$d(x_3, c_1) = \sqrt{((3,49 - 1,3)^2 + (7,01 - 8,04)^2 + (0,92 - 1,02)^2 + (0,18 - 0,18)^2)}$$

$$= 2,42$$

Keterangan: $x_{1,2,3,...}$ adalah kabupaten/kota dan $C_{1,2,3}$ adalah *centroid*.

Perhitungan lengkap x_n, c_1 dapat dilihat pada lampiran 3.

4. Hasil dari perhitungan jarak *eulclidean* dapat menentukan pengelompokkan *cluster* dengan melihat jarak terdekat, dimana ini merupakan langkah selanjutnya dalam metode *k-means*.

Tabel 4.9 : Hasil Perhitungan Jarak *Euclidean K-Means*

C_1	C_2	C_3	JARAK TERDEKAT	CLUSTER
7,20	8,24	6,58	6,58	3
6,23	4,41	2,67	2,67	3
2,42	0,00	6,27	0,00	2
...	...	...	...	...
...	...	...	...	...
5,82	3,46	7,02	3,46	2
6,38	4,08	4,71	4,08	2
7,23	7,86	5,31	5,31	3

Tabel lengkap dapat dilihat pada lampiran 3

5. Menentukan *centroid* baru untuk perhitungan iterasi selanjutnya.

$$C_{11} = (1,03 + 1,3 + 1,23 + 0,84 + 0,45 + 1,03)/6$$

$$= 0,98$$

$$C_{12} = (8,54 + 8,04 + 7,47 + 8,69 + 7,54 + 11,66+)/6$$

$$= 8,66$$

$$C_{13} = (0,94 + 1,02 + 0,93 + 0,99 + 0,49 + 1,87)/6$$

$$= 1,04$$

$$C_{14} = (0,18 + 0,18 + 0,17 + 0,19 + 0,07 + 0,49)/6$$

$$= 0,21$$

Perhitungan dilakukan hingga C_{34}

Tabel 4.10 : *Centroid* Baru

<i>Centroid</i>	TPT	PPM	P_1	P_2
C_1	0,98	8,66	1,04	0,21
C_2	5,33	7,34	0,96	0,22
C_3	5,36	12,87	1,76	0,41

6. Menentukan jarak pada iterasi selanjutnya dengan mengulang langkah ke-3 hingga ke-5 sampai *cluster* yang terbentuk tidak berubah. Berikut hasil akhir dari pengelompokkan yang berhenti pada iterasi ke-5 dengan melihat kesamaannya pada iterasi ke-4.

Tabel 4.11 : Hasil *Clustering* Metode *K-Means* Iterasi ke-4

C_1	C_2	C_3	JT	<i>CLUSTER</i>
6,72	8,01	2,17	2,17	3
5,99	1,30	9,52	1,30	2
2,49	3,14	10,27	2,49	1
7,02	3,59	7,54	3,59	2
0,48	5,35	8,93	0,48	1
0,50	5,08	9,35	0,50	1
4,54	0,52	9,73	0,52	2
4,65	0,49	9,64	0,49	2
3,91	1,10	9,71	1,10	2
1,07	5,20	9,93	1,07	1
1,26	3,75	9,26	1,26	1
8,77	5,37	15,01	5,37	2
5,06	1,54	8,67	1,54	2
8,40	8,87	1,37	1,37	3
0,68	5,55	8,82	0,68	1
1,53	6,00	10,09	1,53	1
3,31	6,39	5,83	3,31	1
3,64	1,63	10,05	1,63	2
5,27	3,20	6,56	3,20	2
2,94	2,03	8,59	2,03	2
4,32	0,77	9,75	0,77	2
1,99	2,95	9,22	1,99	1
3,40	1,75	8,39	1,75	2
13,42	14,15	4,64	4,64	3
14,37	15,63	6,01	6,01	3
6,07	3,23	7,00	3,23	2
4,77	4,41	5,29	4,41	2
7,24	2,47	11,54	2,47	2
4,87	1,30	8,42	1,30	2
7,20	2,33	10,93	2,33	2
5,94	3,53	12,92	3,53	2
6,30	1,86	11,42	1,86	2
6,73	7,14	2,53	2,53	3

Tabel 4.12 : Hasil *Clustering* Metode *K-Means* Iterasi ke-5

C_1	C_2	C_3	JT	<i>CLUSTER</i>
6,86	7,81	3,11	3,11	3
5,78	1,14	10,55	1,14	2
2,22	3,30	11,30	2,22	1
6,89	3,32	8,52	3,32	2
0,72	5,40	9,86	0,72	1
0,51	5,15	10,30	0,51	1
4,31	0,71	10,79	0,71	2
4,42	0,61	10,70	0,61	2
3,67	1,26	10,76	1,26	2
1,00	5,30	10,88	1,00	1
0,99	3,83	10,26	0,99	1
8,49	5,59	16,06	5,59	2
4,88	1,36	9,71	1,36	2
8,51	8,63	2,15	2,15	3
0,95	5,59	9,74	0,95	1
1,59	6,09	11,01	1,59	1
3,51	6,28	6,71	3,51	1
3,38	1,83	11,10	1,83	2
5,19	2,94	7,61	2,94	2
2,75	2,02	9,64	2,02	2
4,09	0,96	10,80	0,96	2
1,74	3,04	10,25	1,74	1
3,22	1,68	9,45	1,68	2
13,57	13,90	3,62	3,62	3
14,55	15,39	4,97	4,97	3
5,96	2,95	8,01	2,95	2
4,77	4,19	6,35	4,19	2
6,99	2,55	12,57	2,55	2
4,70	1,04	9,47	1,04	2
6,96	2,32	11,95	2,32	2
5,67	3,81	13,97	3,81	2
6,04	2,05	12,47	2,05	2
6,82	6,91	3,58	3,58	3

Karena proses pengelompokkan berhenti di iterasi ke-5 maka didapat hasil dari *clustering* tersebut dengan *cluster* 1 sebanyak 9 kabupaten/kota, *cluster* 2 sebanyak 19 kabupaten/kota dan *cluster* 3 sebanyak 5 kabupaten/kota.

Tabel 4.13 : Pengelompokkan Kabupaten/Kota Metode *K-Means*

<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>
Tapanuli Selatan, Tapanuli Utara, Dairi, Humbang, sundutan, Pakpak Samosir, Labuhanbatu Selatan	Mandailing Natal, Tapanuli Tengah, Labuhanbatu, Asahan, Simalungun, Deli Serdang, Langkat, Serdang Bedagai, Batu Bara, Padang Lawas Utara, Padang Lawas, Labuhanbatu Utara, Sibolga, Tanjung Balai, Pematang Siantar, Tebing Tinggi, Medan, Binjai, dan Padang Sidimpuan	Nias, Nias Selatan, Nias Utara, Nias Barat, dan Gunung Sitoli

4.1.6 Validasi *Clustering*

4.1.6.1 Validasi *Agglomerative Hierarchical Clustering* (AHC)

Validasi yang digunakan ialah validasi *Davies Bouldin Index* (DBI) dimana dalam menentukan nilai DBI harus diketahui terlebih dahulu nilai dari rasio perbandingan antar *cluster* ke-*i* dan *cluster* ke-*j* yang dilambangkan dengan $R_{i,j}$ (Persamaan 2.9). Selain itu juga perlu diketahui nilai SSB (*Sum of Square Between*) (Persamaan 2.10) dan nilai SSW (*Sum of Square Within*) (Persamaan 2.11).

Pertama, akan ditentukan jarak data *cluster* menggunakan hasil *clustering* yang diperoleh pada proses *clustering* sebelumnya.

Tabel 4.14 : *Centroid* AHC

<i>Centroid</i>	TPT	PPM	P_1	P_2
C_1	0	3,66	2,1	0
C_2	3,66	0	4,66	0
C_3	1,92	4,66	0	0

$$\begin{aligned}
 d(x_1, c_1) &= \sqrt{((0 - 2,31)^2 + (3,66 - 15,1)^2 + (2,1 - 1,95)^2 + (0 - 0,4)^2)} \\
 &= 11,68
 \end{aligned}$$

$$\begin{aligned}
 d(x_2, c_1) &= \sqrt{((0 - 7,45)^2 + (3,66 - 8,86)^2 + (2,1 - 1,48)^2 + (0 - 0,44)^2)} \\
 &= 11,72
 \end{aligned}$$

$$\begin{aligned}
 d(x_3, c_2) &= \sqrt{((3,66 - 3,49)^2 + (0 - 7,01)^2 + (4,66 - 0,92)^2 + (0 - 0,18)^2)} \\
 &= 10,16
 \end{aligned}$$

$$\begin{aligned}
 d(x_4, c_2) &= \sqrt{((3,66 - 7,81)^2 + (0 - 11,5)^2 + (4,66 - 1,58)^2 + (0 - 0,3)^2)} \\
 &= 7,95
 \end{aligned}$$

Perhitungan dilakukan hingga $d(x_i, c_j)$ dimana i adalah kabupaten/kota ke- i dan j adalah klaster. Setelah perhitungan jarak antar *cluster* dilakukan, maka selanjutnya ialah menghitung nilai SSW menggunakan rumus pada persamaan 2.11

$$SSW_1 = (11,68 + 11,72)/2$$

$$= 11,70$$

$$\begin{aligned}
 SSW_2 &= (10,16 + 7,95 + 12,61 + 9,68 + 9,14 + 8,95 + 9,41 + 8,85 + \\
 &\quad 8,70 + 8,78 + 7,42 + 9,85 + 9,20 + 12,28 + 8,45 + 12,02 + \\
 &\quad 9,50 + 8,78 + 8,88 + 9,80 + 12,26 + 12,68 + 9,51 + 10,34 \\
 &\quad + 6,81 + 8,75 + 9,94)/27 \\
 &= 9,66
 \end{aligned}$$

$$SSW_3 = (12,25 + 17,34 + 18,31)/3$$

$$= 15,96$$

Langkah ketiga ialah menghitung nilai SSB menggunakan rumus pada persamaan 2.10.

$$\begin{aligned}
 SSB_{1,C_1} &= \sqrt{((0 - 0)^2 + (3,66 - 3,66)^2 + (2,1 - 2,1)^2 + (0 - 0)^2)} \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
 SSB_{1,C_2} &= \sqrt{((0 - 3,66)^2 + (3,66 - 0)^2 + (2,1 - 4,66)^2 + (0 - 0)^2)} \\
 &= 5,77
 \end{aligned}$$

$$\begin{aligned}
SSB_{1,C_3} &= \sqrt{((0 - 1,92)^2 + (3,66 - 4,66)^2 + (2,1 - 0)^2 + (0 - 0)^2)} \\
&= 3,02 \\
SSB_{2,C_1} &= \sqrt{((3,66 - 0)^2 + (0 - 3,66)^2 + (4,66 - 2,1)^2 + (0 - 0)^2)} \\
&= 5,77 \\
SSB_{2,C_2} &= \sqrt{((3,66 - 3,66)^2 + (0 - 0)^2 + (4,66 - 4,66)^2 + (0 - 0)^2)} \\
&= 0 \\
SSB_{2,C_3} &= \sqrt{((3,66 - 1,92)^2 + (0 - 4,66)^2 + (4,66 - 0)^2 + (0 - 0)^2)} \\
&= 6,82 \\
SSB_{3,C_1} &= \sqrt{((1,92 - 0)^2 + (4,66 - 3,66)^2 + (0 - 2,1)^2 + (0 - 0)^2)} \\
&= 3,02 \\
SSB_{3,C_2} &= \sqrt{((1,92 - 3,66)^2 + (4,66 - 0)^2 + (0 - 4,66)^2 + (0 - 0)^2)} \\
&= 6,82 \\
SSB_{3,C_3} &= \sqrt{((1,92 - 1,92)^2 + (4,66 - 4,66)^2 + (0 - 0)^2 + (0 - 0)^2)} \\
&= 0
\end{aligned}$$

Tabel 4.15 : Matriks SSB Metode AHC

SSB	CENTROID		
	1	2	3
1	0	5,77	3,02
2	5,77	0	6,82
3	3,02	6,82	0

Langkah keempat dalam mencari nilai DBI yaitu dengan mencari nilai rasio perbandingan antara *cluster* ke-*i* dan *cluster* ke-*j* dengan menggunakan persamaan 2.9.

$$R_{1,2} = \frac{SSW_1 + SSW_2}{SSB_{1,2}} = \frac{11,70 + 9,66}{5,77} = 3,70$$

$$R_{1,3} = \frac{SSW_1 + SSW_3}{SSB_{1,3}} = \frac{11,70 + 15,96}{3,02} = 9,17$$

$$R_{2,3} = \frac{SSW_2 + SSW_3}{SSB_{2,3}} = \frac{9,66 + 15,96}{6,82} = 3,76$$

Jika ditransformasikan kedalam bentuk matriks maka akan seperti tabel beserta dengan R_{max} yang dibutuhkan dalam perhitungan akhir mencari nilai DBI.

Tabel 4.16 : R_{max} Metode AHC

R	1	2	3	R_{max}
1	0	3,70	9,17	9,17
2	3,70	0	3,76	3,76
3	9,17	3,76	0	9,17

Untuk langkah terakhir perhitungan DBI digunakan persamaan 2.8 yaitu:

$$DBI = \frac{9,17+3,76+9,17}{3} = 7,37$$

4.1.6.2 Validasi *K-Means Clustering*

Validasi pada *K-Means Clustering* sama halnya dengan validasi pada AHC yaitu dalam menentukan nilai DBI harus diketahui terlebih dahulu nilai dari rasio perbandingan antar *cluster* ke-*i* dan *cluster* ke-*j* yang dilambangkan dengan $R_{i,j}$ (Persamaan 2.9) . Selain itu juga perlu diketahui nilai SSB (*Sum of Square Between*) (Persamaan 2.10) dan nilai SSW (*Sum of Square Within*) (Persamaan 2.11).

Pertama, akan ditentukan adalah jarak data *cluster* menggunakan hasil *clustering* yang diperoleh pada proses *clustering* sebelumnya serta *centroid* yang merupakan hasil dari *clustering*.

Tabel 4.17 : *Centroid K-Means*

<i>Centroid</i>	TPT	PPM	P_1	P_2
C_1	1,71	8,33	1,03	0,21
C_2	6,42	8,51	1,19	0,28
C_3	2,57	18,17	2,33	0,53

$$\begin{aligned} d(x_3, c_1) &= \sqrt{((1,71 - 2,31)^2 + (8,33 - 15,1)^2 + (1,03 - 1,95)^2 + (0,21 - 0,4)^2)} \\ &= 2,22 \end{aligned}$$

$$\begin{aligned} d(x_5, c_1) &= \sqrt{((1,71 - 3,49)^2 + (8,33 - 7,01)^2 + (1,03 - 0,92)^2 + (0,21 - 0,18)^2)} \\ &= 0,72 \end{aligned}$$

$$\begin{aligned} d(x_6, c_1) &= \sqrt{((1,71 - 1,03)^2 + (8,33 - 8,54)^2 + (1,03 - 0,94)^2 + (0,21 - 0,18)^2)} \\ &= 0,50 \end{aligned}$$

Perhitungan dilakukan hingga $d(x_i, c_j)$ dimana i adalah kabupaten/kota ke- i dan j adalah klaster. Setelah perhitungan jarak antar *cluster* dilakukan, maka selanjutnya ialah menghitung nilai SSW menggunakan rumus pada persamaan 2.11.

$$\begin{aligned} SSW_1 &= (2,22 + 0,72 + 0,5 + 0,99 + 1 + 0,94 + 1,59 + 3,51 + 1,74)/9 \\ &= 1,47 \end{aligned}$$

$$\begin{aligned} SSW_2 &= (2,02 + 0,96 + 1,68 + 2,95 + 4,19 + 2,55 + 1,04 + 3,81 \\ &\quad + 2,05 + 1,14 + 3,32 + 0,71 + 0,62 + 1,26 + 5,6 + 1,36 \\ &\quad + 1,83 + 2,94 + 2,32)/19 \\ &= 2,23 \end{aligned}$$

$$\begin{aligned} SSW_3 &= (3,11 + 2,14 + 3,63 + 4,98 + 3,57)/5 \\ &= 3,49 \end{aligned}$$

Langkah ketiga ialah menghitung nilai SSB menggunakan rumus pada persamaan 2.10.

$$\begin{aligned} SSB_{1,C_1} &= \sqrt{((1,71 - 1,71)^2 + (8,33 - 8,33)^2 + (1,03 - 1,03)^2 + (0,21 - 0,21)^2)} \\ &= 0 \end{aligned}$$

$$\begin{aligned} SSB_{1,C_2} &= \sqrt{((1,71 - 6,42)^2 + (8,33 - 8,51)^2 + (1,03 - 1,19)^2 + (0,21 - 0,28)^2)} \\ &= 4,72 \end{aligned}$$

$$\begin{aligned} SSB_{1,C_3} &= \sqrt{((1,71 - 2,57)^2 + (8,33 - 18,17)^2 + (1,03 - 2,33)^2 + (0,21 - 0,53)^2)} \\ &= 9,97 \end{aligned}$$

Perhitungan dilakukan hingga SSB_{3,C_3}

Tabel 4.18 : Matriks SSB Metode *K-Means*

SSB	CENTROID		
	1	2	3
1	0	4,72	9,97
2	4,72	0	10,46
3	9,97	10,46	0

Langkah keempat dalam mencari nilai DBI yaitu dengan mencari nilai rasio perbandingan antara *cluster* ke-*i* dan *cluster* ke-*j* dengan menggunakan persamaan 2.9.

$$R_{1,2} = \frac{SSW_1 + SSW_2}{SSB_{1,2}} = \frac{1,47 + 2,23}{4,72} = 0,78$$

$$R_{1,3} = \frac{SSW_1 + SSW_3}{SSB_{1,3}} = \frac{1,47 + 3,49}{9,97} = 0,50$$

$$R_{2,3} = \frac{SSW_2 + SSW_3}{SSB_{2,3}} = \frac{2,23 + 3,49}{10,46} = 0,55$$

Jika ditransformasikan kedalam bentuk matriks maka akan seperti tabel beserta dengan R_{max} yang dibutuhkan dalam perhitungan akhir mencari nilai DBI.

Tabel 4.19 : R_{max} *K-Means*

R	1	2	3	R_{max}
1	0	0,78	0,50	0,78
2	0,78	0	0,55	0,78
3	0,50	0,55	0	0,55

Untuk langkah terakhir perhitungan DBI digunakan persamaan 2.8 yaitu:

$$DBI = \frac{0,78 + 0,78 + 0,55}{3} = 0,70$$

4.2 Pembahasan

Berdasarkan uraian dari hasil penelitian didapat *cluster* dengan tingkat kemiskinan rendah, sedang dan tinggi menggunakan dua metode, yaitu Algoritma *K-Means Clustering* dan Algoritma *Agglomerative Hierarchical Clustering*.

Hasil *clustering* menggunakan metode *agglomerative hierarchical clustering* (AHC), pada *cluster* 1 diperoleh 2 kabupaten/kota yang terdiri dari Nias, dan Gunung Sitoli. Pada *cluster* 2 diperoleh 28 kabupaten/kota yaitu Mandailing Natal, Tapanuli Selatan, Karo, Labuhanbatu Selatan, Padang Lawas Utara, Labuhanbatu Utara, Tapanuli Utara, Humbang Hasundutan, Toba, Simalungun, Dairi, Serdang Bedagai, Labuhanbatu, Padang Lawas, Asahan, Langkat, Pakpak, Barat, Tebing Tinggi, Pematang Siantar, Medan, Padang Sidempuan, Tapanuli Tengah, Batu Bara, Sibolga, Tanjung Balai, Binjai, Deli Serdang dan Samosir serta pada *cluster* 3 diperoleh 3 kabupaten/kota yaitu Nias Selatan, Nias Utara, dan Nias Barat.

Hasil *clustering* yang diperoleh menggunakan metode k-means berbeda dengan metode AHC yaitu pada *cluster* 1 diperoleh 9 kabupaten/kota dengan tingkat kemiskinan rendah, pada *cluster* 2 diperoleh 19 kabupaten/kota dengan tingkat kemiskinan sedang dan pada *cluster* 3 diperoleh 5 kabupaten/kota dengan tingkat kemiskinan tinggi. Untuk *cluster* 1 terdiri dari Tapanuli Selatan, Tapanuli Utara, Toba, Dairi, Karo, Humbang Hasundutan, Pakpak Barat, Samosir dan Labuhanbatu Selatan. Pada *cluster* 2 terdiri dari Mandailing Natal, Tapanuli Tengah, Labuhanbatu Asahan, Simalungun, Deli Serdang, Langkat, Serdang Bedagai, Batu Bara, Padang Lawas Utara, Padang Lawas, Labuhanbatu Utara, Sibolga, Tanjung Balai, Pematang Siantar, Tebing Tinggi, Medan, Binjai dan Padang Sidimpuan, sedangkan pada *cluster* 3 terdiri dari 5 kabupaten/kota yaitu Nias, Nias Selatan, Nias Utara, Nias Barat, dan Gunung Sitoli.

Pada penelitian ini digunakan dua metode dalam mengelompokkan kabupaten/kota berdasarkan tingkat kemiskinan rendah, sedang dan tinggi. Oleh karena itu perlu dilakukan validasi *clustering* untuk melihat metode mana yang menghasilkan *clustering* terbaik menggunakan validasi *Davies Bouldin Index* (DBI). Berdasarkan uraian hasil penelitian pada bab iv, nilai DBI untuk metode *agglomerative hierarchical clustering* ialah sebesar 7,37 dan untuk metode *k-means clustering* sebesar 0,70. Sehingga dalam proses *clustering* metode yang terbaik ialah menggunakan k-means *clustering* karena nilai DBI yang diperoleh lebih kecil.