

BAB 2

TINJAUAN PUSTAKA

2.1 Statistika Deskriptif

Statistika deskriptif adalah metode yang berkaitan dengan pengumpulan dan penyajian suatu gugus data sehingga memberikan informasi yang berguna. Statistik deskriptif memberikan informasi hanya mengenai data yang dipunyai, contohnya dalam bentuk tabel, grafik, dan diagram (Walpole, 1988). Statistika adalah suatu pengetahuan tentang data. Ini melibatkan pengumpulan, klasifikasi, merangkum, pengorganisasian, analisis, dan interpretasi data (Ramachandran & Tsokos, 2009).

Data adalah hasil observasi, pengamatan, pengukuran, atau penelitian yang telah dikumpulkan (Hair, dkk, 2010). Dalam suatu penelitian, data memiliki peran yang sangat penting. Data dapat dikelompokkan menjadi dua berdasarkan sifatnya, yaitu data kuantitatif dan data kualitatif (Hair, dkk, 2010). Data kuantitatif adalah observasi yang diukur dalam skala numerik. Data non-numerik yang hanya dapat digolongkan ke dalam salah satu kelompok kategori dikatakan data kualitatif atau kategorikal (Ramachandran & Tsokos, 2009). Menurut Walpole (1988), data kuantitatif adalah data hasil cacahan ataupun hasil pengukuran, atau mungkin dengan data kualitatif yang di klasifikasikan menurut kriteria tertentu.

Sumber pengetahuan statistik terletak pada data. Setelah nilai data sampel diperoleh, salah satu cara untuk mengenal data tersebut adalah dengan menampilkannya dalam tabel atau grafik. Bagan dan grafik merupakan alat yang sangat penting dalam statistik karena keduanya mengkomunikasikan informasi secara visual (Ramachandran & Tsokos, 2009). Tampilan visual ini dapat mengungkap pola perilaku variabel yang diteliti. Tampilan grafis yang paling umum adalah tabel frekuensi, diagram lingkaran, grafik batang, diagram Pareto, dan histogram (Ramachandran & Tsokos, 2009).

Misalkan terdapat sampel dengan nilai $X_1, X_2, \dots, X_n$. Terdapat banyak karakteristik yang terkait dengan kumpulan data ini, misalnya kecenderungan sentral dan variabilitas. Ukuran tendensi sentral diberikan oleh rata-rata, median, atau modus

sampel. Sedangkan ukuran variabilitas biasanya diberikan oleh varians sampel, simpangan baku, atau rentang antarkuartil (Ramachandran & Tsokos, 2009).

Menurut Ramachandra dan Tsokos (2009), rumus yang digunakan untuk menghitung rata-rata sampel dan varians sampel adalah sebagai berikut.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (2.1)$$

$$s^2 = \frac{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2}{n(n-1)} \quad (2.2)$$

Keterangan:

s^2 : Nilai Varians

X_i : Objek, $i = 1, 2, \dots, n$.

$\bar{X}$: Rata-rata objek

n : Jumlah objek

Varians sampel (s^2) adalah ukuran variabilitas atau penyebaran data di sekitar rata-rata sampel ($\bar{X}$). Semakin besar nilai varians, maka semakin besar penyebarannya (Ramachandran & Tsokos, 2009).

2.2 Standarisasi Data

Standarisasi data memiliki tujuan utama yaitu mengubah nilai-nilai variabel awal menjadi nilai-nilai baru dengan rata-rata nol dan simpangan baku satu. Standarisasi data dilakukan jika satuan dalam data mempunyai perbedaan yang besar. Jika data yang digunakan mempunyai satuan yang berbeda secara signifikan pada setiap variabel yang akan dijadikan sebagai objek penelitian, maka data tersebut perlu dilakukan standarisasi dengan mengubah data yang ada ke nilai *Z-score* (Lind, dkk, 2007). Persamaan untuk standarisasi adalah sebagai berikut.

$$Z = \frac{X_i - \bar{X}}{s} \quad (2.3)$$

Keterangan:

Z : Nilai *Z-score*

X_i : Objek, $i = 1, 2, \dots, n$.

$\bar{X}$: Rata-rata objek

s : Simpangan baku objek

Rumus untuk menghitung simpangan baku adalah sebagai berikut.

$$s = \sqrt{s^2} \quad (2.4)$$

Keterangan:

s : Simpangan baku objek

s^2 : Nilai Varians

2.3 Uji Multikolinearitas

Multikolinearitas terjadi pada model regresi dengan lebih dari satu variabel independen (regresi berganda), dimana terjadi korelasi yang kuat antar variabel independen. Multikolinearitas dapat dilihat dari nilai VIF (*Variance Inflation Factor*). Rumus VIF adalah sebagai berikut (Lind, dkk, 2007).

$$VIF = \frac{1}{(1 - R^2)} \quad (2.5)$$

Keterangan:

VIF : *Variance Inflation Factor*

R^2 : Koefisien determinasi

Jika nilai $VIF > 10$ maka dapat dikatakan terjadi multikolinearitas antar variabel independen (Lind, dkk, 2006). Rumus untuk mencari nilai koefisien determinasi adalah sebagai berikut.

$$R^2 = \frac{\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \quad (2.6)$$

Keterangan:

R^2 : Koefisien determinasi

$\boldsymbol{\beta}$: Vektor dari parameter

$\mathbf{X}$: Matriks dari variabel independen

$\mathbf{y}$: Vektor dari variabel dependen

Untuk menghitung nilai dari parameter-parameter yang tidak diketahui nilainya digunakan rumus:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (2.7)$$

Dimana nilai invers dari matriks $\mathbf{X}^T\mathbf{X}$ dapat dihitung menggunakan Operasi Baris Elementer (OBE). Terdapat tiga acara untuk melakukan perhitungan menggunakan OBE adalah sebagai berikut (Anton, 2005).

1. Kalikan suatu baris dengan konstanta tidak nol
2. Tukarkan dua buah baris
3. Tambahkan suatu baris dengan kelipatan suatu baris lainnya

Persamaan yang digunakan untuk menghitung $(\mathbf{X}^T\mathbf{X})^{-1}$ adalah sebagai berikut.

$$[\mathbf{X}^T\mathbf{X} \mid \mathbf{I}] \rightarrow [\mathbf{I} \mid (\mathbf{X}^T\mathbf{X})^{-1}] \quad (2.8)$$

Keterangan:

$\mathbf{I}$: Matriks identitas

Matriks identitas adalah sebuah matriks persegi dengan bilangan 1 pada diagonal utamanya dan 0 pada entri-entri lainnya. Contoh matriks identitas adalah sebagai berikut (Anton, 2005).

$$\mathbf{I}_{(2 \times 2)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{I}_{(3 \times 3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{I}_{(4 \times 4)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

2.4 Analisis Cluster

Analisis *cluster* merupakan salah satu analisis multivariat yang bertujuan untuk mengelompokkan objek-objek berdasarkan kemiripan karakteristik yang dimilikinya. Tingkat kemiripan karakteristik yang dimaksud adalah objek-objek dalam suatu kelompok sangat tinggi sedangkan tingkat kemiripan karakteristik objek antar kelompok satu dengan yang lainnya rendah (Hair, dkk, 2010).

Prosedur dalam pembentukan kelompok terbagi menjadi 2, yaitu dengan hierarki dan non-hierarki. Metode non-hierarki adalah pengelompokan objek dimana banyaknya kelompok yang akan dibentuk dapat ditentukan terlebih dahulu (Johnson & Wichern, 2007). Sedangkan metode hierarki merupakan pengelompokkan objek dimana banyak kelompok yang akan terbentuk belum diketahui jumlahnya (Johnson, 1967).

Dalam metode algoritma setiap objek pengamatan pada mulanya dianggap sebagai kelompok masing-masing, sehingga terdapat kelompok sebanyak jumlah objek pengamatan. Kemudian dua kelompok yang memiliki kesamaan terdekat

bergabung menjadi satu kelompok baru, sehingga jumlah kelompok akan berkurang satu setiap tahap pengelompokkan (Hardle & Simar, 2007). Untuk Teknik algoritma ini terdiri dari tiga metode, yaitu metode *Linkage*, metode *Variance* dan metode *Centroid*. Dalam metode *Linkage* terdiri dari metode *Single Linkage*, *Complete Linkage*, dan *Average Linkage*. Sedangkan untuk metode *Variance* terdiri dari metode *Ward* (Hardle & Simar, 2007).

2.4.1 Perhitungan Jarak

Tujuan utama dalam analisis *cluster* yaitu mengidentifikasi vektor dari objek pengamatan yang serupa dan mengelompokkannya menjadi kelompok baru. Terdapat beberapa teknik yang biasa digunakan untuk menghitung jarak antar setiap pasangan objek. Jarak antara dua objek merupakan sebuah ukuran kedekatan yang tepat (Rencher, 2002). Metode yang digunakan untuk menghitung jarak dalam analisis kelompok adalah sebagai berikut.

A. Jarak *Euclidean*

Jarak *Euclidean* adalah jarak antar kelompok dari p variabel, jarak *Euclidean* merupakan perhitungan jarak antar objek yang paling sering digunakan. Ukuran satuan yang digunakan dalam melakukan pengukuran dari variabel-variabel dengan jarak *Euclidean* merupakan pertimbangan yang penting. Jika antar variabelnya memiliki perbedaan satuan yang cukup besar, untuk mengatasi masalah perbedaan satuan ini dapat dilakukan standarisasi (Rencher, 2002).

Jarak *Euclidean* antara kelompok ke- i dan ke- j dari p variabel dirumuskan sebagai berikut (Rencher, 2002).

$$d_{(i,j)} = \sqrt{\left[\sum_{k=1}^p (X_{ik} - X_{jk})^2 \right]} \quad (2.9)$$

Keterangan:

$d_{(i,j)}$: Jarak antar objek i ke objek j

X_{ik} : Objek i pada variabel ke- k

X_{jk} : Objek j pada variabel ke- k

p : Jumlah variabel

Dengan $k = 1, 2, \dots, p$ dan $i, j = 1, 2, \dots, n$.

B. Jarak *Mahalanobis*

Jarak *Mahalanobis* digunakan untuk mengatasi masalah perbedaan skala yang terdapat pada data, selain itu juga mempertimbangkan pengaruh korelasi antar variabel. Jika terdapat korelasi atau multikolinearitas antar variabel, alternatif jarak yang digunakan adalah jarak *Mahalanobis* (Everitt, dkk, 2011).

Jarak *Mahalanobis* dapat didefinisikan yaitu jarak antar dua titik yang melibatkan kovarians atau korelasi yang ada diantara variabel. Menurut Everitt, dkk, (2011), jarak *Mahalanobis* dari dua objek dapat dinyatakan dalam bentuk vektor dan matriks. Rumus untuk jarak *Mahalanobis* adalah sebagai berikut.

$$d_{(i,j)} = \sqrt{(X_i - X_j)^T W^{-1} (X_i - X_j)} \quad (2.10)$$

Keterangan:

$d_{(i,j)}$: Jarak antar objek i ke objek j

X_i : Objek ke- i , dimana $i = 1, 2, \dots, n$.

X_j : Objek ke- j , dimana $j = 1, 2, \dots, n$.

W : Matriks kovarians dari objek

Penggunaan invers matriks kovarians dalam jarak *Mahalanobis* memiliki manfaat untuk menstandarkan semua variabel ke varians yang sama dan menghilangkan korelasi (Everitt, dkk, 2011).

2.4.2 Metode *Centroid*

Metode *centroid* merupakan salah satu metode pengelompokan hierarki. Metode ini dilakukan dengan cara menggabungkan dua kelompok melalui jarak terdekat pada titik pusat antar kelompok (Rencher, 2002). Pada metode *centroid*, jarak antara dua kelompok A dan B didefinisikan sebagai berikut.

$$d(\bar{X}_A, \bar{X}_B) = \bar{X}_{AB} \quad (2.11)$$

Keterangan:

$\bar{X}_A$: *Centroid* untuk objek di A

$\bar{X}_B$: *Centroid* untuk objek di B

$d(\bar{X}_A, \bar{X}_B)$: Jarak antar *centroid* kelompok A dan kelompok B

$\bar{X}_{AB}$: *Centroid* untuk objek di AB

$\bar{X}_A$ dan $\bar{X}_B$ dihitung dengan cara biasa, yaitu $\bar{X}_A = \sum_{i=1}^{n_A} \frac{X_i}{n_A}$ dan $\bar{X}_B = \sum_{j=1}^{n_B} \frac{X_j}{n_B}$. Dua kelompok dengan jarak terkecil antar *centroid* digabungkan. Setelah dua kelompok A dan B digabungkan, pusat massa kelompok AB baru diberikan rata-rata tertimbang (Rencher, 2002).

$$\bar{X}_{AB} = \frac{n_A \bar{X}_A + n_B \bar{X}_B}{n_A + n_B} \quad (2.12)$$

$$\bar{X}_{AB} = \frac{n_A \left(\sum_{i=1}^{n_A} \frac{X_i}{n_A} \right) + n_B \left(\sum_{j=1}^{n_B} \frac{X_j}{n_B} \right)}{n_A + n_B} \quad (2.13)$$

Keterangan:

X_i : Objek ke- i , $i = 1, 2, \dots, n$.

X_j : Objek ke- j , $j = 1, 2, \dots, n$.

$\bar{X}_{AB}$: *Centroid* untuk objek di AB

$\bar{X}_A$: *Centroid* untuk objek di A

$\bar{X}_B$: *Centroid* untuk objek di B

n_A : Jumlah objek di A

n_B : Jumlah objek di B

Keunggulan dari metode *centroid* adalah memiliki tingkat ketepatan hasil terbaik untuk pengelompokan dan metode ini juga baik untuk data yang mengandung *outlier* (Rencher, 2002). Metode *centroid* memiliki perbedaan dengan metode *Linkage* yang beroperasi secara langsung pada matriks jarak, metode pengelompokan *centroid* memerlukan akses ke data asli (Everitt, dkk, 2011).

Berikut langkah-langkah pengelompokan dengan menggunakan metode *centroid* (Rencher, 2002).

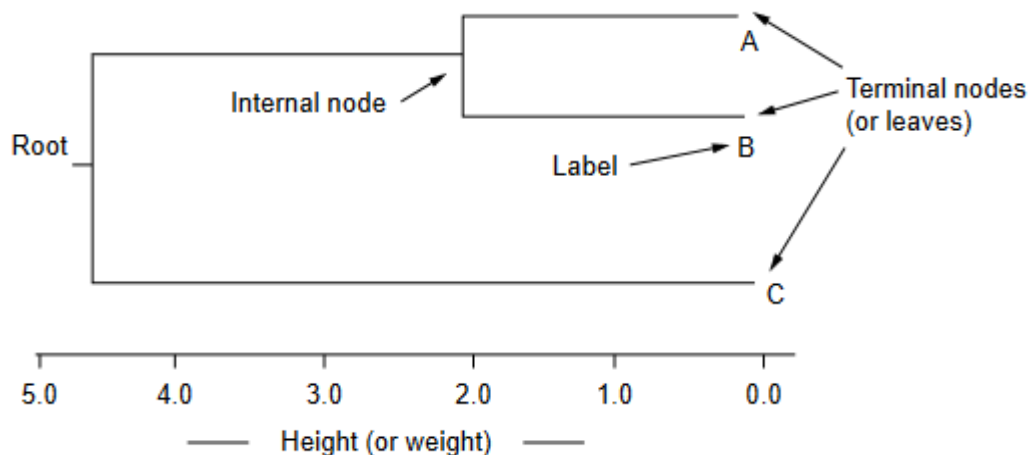
- Mulai dengan membuat kelompok sebanyak n objek. Dalam sebuah kelompok terdiri dari sebuah objek.
- Membuat matriks jarak dengan menghitung jarak antar kelompok. Perhitungan jarak menggunakan jarak *Euclidean* pada Persamaan (2.9).
- Mencari sepasang kelompok yang memiliki jarak terdekat pada matriks jarak.

- d. Menghitung vektor rata-rata (*centroid*) antar kelompok yang memiliki jarak terdekat menggunakan Persamaan (2.13). Perhitungan *centroid* dilakukan pada data asli yang telah distandarisasi.
- e. Langkah (b) – (d) diulang sampai seluruh objek menjadi satu kelompok.
- f. Memperoleh hasil pengelompokan.

2.4.3 Dendrogram

Pada metode algoritma, hasil dari analisis pengelompokan yang menggunakan metode hierarki dapat diwakili oleh diagram dua dimensi yang dikenal sebagai dendrogram. Dendrogram menggambarkan bagaimana objek-objek digabungkan berdasarkan tingkat kesamaan atau jarak di antara objek-objek tersebut (Rencher, 2002). Dendrogram mengilustrasikan prosesnya, dan partisi yang dihasilkan pada setiap tahap. Ketinggian dalam diagram ini mewakili jarak di mana setiap pengelompokan yang terjadi (Everitt, dkk, 2011).

Menurut Everitt, dkk. (2011) berikut adalah gambaran dendrogram beserta komponen-komponen penyusunnya.



Gambar 2. 1. Dendrogram

(Everitt, dkk, 2011)

Penjelasan mengenai komponen-komponen dari dendrogram berdasarkan Gambar 2.1 adalah sebagai berikut.

1. Akar (*Root*): Titik awal dari dendrogram. *Root* menunjukkan pengelompokan seluruh objek saat menjadi satu kelompok besar.
2. *Terminal node*: Titik-titik yang diberi label dengan nama objek yang diwakilinya (A, B, dan C).

3. *Internal node*: Titik-titik percabangan yang menggabungkan kelompok objek berdasarkan jarak terdekat.
 - 1) *Node* pertama menggabungkan A dan B
 - 2) *Node* kedua menggabungkan (A, B) dan C
4. Tinggi (*Height*): Tinggi *node* menunjukkan jarak antar kelompok yang digabungkan pada setiap *internal node*. Berdasarkan Gambar 2.1, contoh tinggi *node* adalah sebagai berikut.
 - 1) Tinggi *node* penggabungan A dan B: *Node* yang menggabungkan A dan B berada pada angka 2, maka jarak antara A dan B adalah 2.
 - 2) Tinggi *node* penggabungan (A, B) dan C: *Node* yang menggabungkan A dan B berada pada angka 4,8, maka jarak antara A dan B adalah 4,8.

Dalam pengelompokan hierarki, kita dapat memilih jumlah kelompok dari dendrogram dengan memotong cabang-cabang pada tingkat jarak tertentu yang digunakan oleh salah satu sumbu. Menentukan jumlah kelompok dilakukan dengan cara mencari jarak terjauh dimana kelompok terbentuk (Rencher, 2002).

Langkah-langkah membuat dendrogram adalah sebagai berikut.

1. Menghitung matriks jarak.
2. Mengelompokkan objek menggunakan metode pengelompokan hierarki, kemudian memperoleh hasil pengelompokan.
3. Membuat titik-titik yang diberi label dengan nama setiap objek yang diwakilinya (*terminal node*).
4. Membuat titik-titik percabangan yang menggabungkan kelompok objek berdasarkan jarak terdekat (*internal node*).
5. Membuat *internal node* sampai membentuk satu kelompok besar, yang menunjukkan pengelompokan seluruh objek.
6. Melihat jarak terjauh pada tinggi dendrogram untuk menentukan pusat pemotongan dendrogram untuk memisahkan kelompok.
7. Melakukan pemotongan dendrogram, dengan cara membuat garis vertikal yang menyeberangi dendrogram pada tinggi yang dipilih. Garis ini menunjukkan titik dimana kelompok dibawahnya dipisahkan menjadi kelompok yang lebih kecil.

8. Memperhatikan hasil pemotongan, jumlah kelompok yang terbentuk dari hasil pemotongan dendrogram dapat digunakan.

2.4.4 Validasi Kelompok

Validasi kelompok digunakan untuk menguji kebaikan/ketepatan hasil kelompok. Salah satu metode validasi data yang dapat digunakan yaitu Koefisien *Silhouette*. Koefisien *Silhouette* adalah ukuran seberapa mirip suatu objek dengan kelompoknya sendiri dibandingkan dengan kelompok lainnya (Everitt, dkk, 2011).

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (2.14)$$

Rumus untuk menghitung nilai a_i dan b_i adalah sebagai berikut.

$$a_i = \frac{1}{C_i - 1} \sum_{j \in C_i, i \neq j} d_{(i,j)} \quad (2.15)$$

$$b_i = \min \frac{1}{C_k} \sum_{j \in C_k} d_{(i,j)} \quad (2.16)$$

Keterangan:

S_i : Koefisien *Silhouette*, $i = 1, 2, \dots, n$.

a_i : Rata-rata jarak antara objek i dengan objek lainnya pada kelompok

b_i : Rata-rata jarak minimum dari objek i dengan kelompok lainnya

C_i : Jumlah objek dalam kelompok C_i

C_k : Jumlah objek dalam kelompok C_k

Untuk mengetahui rata-rata koefisien *silhouette* pada jumlah kelompok tertentu, dapat dilakukan perhitungan rata-rata dengan cara biasa yaitu sebagai berikut.

$$S_c = \frac{1}{n} \sum_{i=1}^n S_i \quad (2.17)$$

Keterangan:

S_i : Koefisien *silhouette*, $i = 1, 2, \dots, n$.

S_c : Rata-rata koefisien *silhouette*

n : Jumlah objek

Rata-rata koefisien *Silhouette* berkisar dari -1 hingga 1 , dimana kelompok semakin baik apabila koefisien *Silhouette* semakin mendekati 1 (Everitt, dkk, 2011). Rata-rata koefisien *Silhouette* -1 menunjukkan bahwa objek-objek dalam kelompok lebih dekat dengan kelompok lain daripada kelompoknya sendiri, artinya hasil pengelompokan tersebut buruk. Rata-rata koefisien *Silhouette* 0 menunjukkan bahwa objek-objek berada di perbatasan antara dua kelompok, artinya pemisahan tidak jelas. Sebaliknya, rata-rata koefisien *Silhouette* 1 menunjukkan bahwa objek-objek sangat cocok dengan kelompoknya sendiri dan jauh dari kelompok lain, artinya hasil pengelompokan sangat baik (Han dkk., 2012). Koefisien *silhouette* memiliki kriteria sebagai berikut.

- a) $0,71 - 1,00$: Struktur yang kuat.
- b) $0,51 - 0,70$: Struktur yang standar.
- c) $0,26 - 0,50$: Struktur yang lemah.
- d) $\leq 0,25$: Struktur yang tidak sesuai.

(Kaufman dan Rousseeuw, 2005)

2.4.5 Contoh Analisis Kelompok

Untuk mengilustrasikan analisis pengelompokan dengan menggunakan metode *centroid*, perhitungan *centroid* akan digunakan pada kumpulan data bivariat pada Tabel 2.1.

Tabel 2. 1 Data untuk contoh metode *centroid*

No	Provinsi	X_1	X_2	X_3
1	Aceh	70,18	14,37	9,44
2	DKI Jakarta	73,32	13,08	11,31
3	Kalimantan Barat	71,02	12,66	7,59
4	Maluku	66,45	14,00	10,19
5	Papua	66,23	11,14	8,69

(BPS, 2023)

Misalkan jarak *Euclidean* dipilih sebagai ukuran jarak antar provinsi, maka perhitungan jarak Provinsi Aceh dengan Provinsi DKI Jakarta adalah sebagai berikut.

$$\begin{aligned}
 d_{(1,2)} &= ((70,18 - 73,32)^2 + (14,37 - 13,08)^2 + (9,44 - 11,31)^2)^{\frac{1}{2}} \\
 &= (9,86 + 1,66 + 3,50)^{\frac{1}{2}} \\
 &= 3,88
 \end{aligned}$$

Dengan cara yang sama dilakukan perhitungan jarak untuk provinsi lainnya, sehingga menghasilkan matriks jarak berukuran 5×5 berikut ini.

Tabel 2. 2 Hasil Perhitungan Jarak *Euclidean* untuk contoh metode *centroid*

Provinsi	Aceh	DKI Jakarta	Kalimantan Barat	Maluku	Papua
Aceh	0,00	3,88	2,66	3,82	5,16
Provinsi	Aceh	DKI Jakarta	Kalimantan Barat	Maluku	Papua
DKI Jakarta	3,88	0,00	4,39	7,02	7,80
Kalimantan Barat	2,66	4,39	0,00	5,43	5,14
Maluku	3,82	7,02	5,43	0,00	3,24
Papua	5,16	7,80	5,14	3,24	0,00

Jarak terkecil adalah 2,66 antara Aceh dan Kalimantan Barat, sehingga kedua provinsi tersebut digabungkan menjadi kelompok pertama. Kelompok yang terbentuk disimbolkan dengan C, maka C1 merupakan kelompok pertama yang beranggotakan Aceh dan Kalimantan Barat.

Perhitungan *centroid* menggunakan data pada Tabel 2.1. Perhitungan nilai *centroid* untuk C1 menggunakan Persamaan (2.13). Perhitungan *centroid* variabel X_1 untuk C1 adalah sebagai berikut.

$$\bar{X}_{1(1,3)} = \frac{(70,18) + (71,02)}{1 + 1} = 70,60$$

Dengan cara yang sama lakukan perhitungan *centroid* C1 untuk variabel lainnya. Maka didapatkan nilai *centroid* C1 yang dapat dilihat pada Tabel 2.3 berikut ini.

Tabel 2. 3 Contoh hasil perhitungan nilai *centroid* C1

No	Provinsi	$\bar{X}_1$	$\bar{X}_2$	$\bar{X}_3$
1 & 3	C1	70,60	13,52	8,52

Selanjutnya jarak kembali dihitung berdasarkan data provinsi DKI Jakarta, Maluku, Papua, dan C1. Perhitungan jarak Provinsi DKI Jakarta dengan C1 adalah sebagai berikut.

$$\begin{aligned}
 d_{(2,C1)} &= ((73,32 - 70,60)^2 + (13,08 - 13,52)^2 + (11,31 - 8,52)^2)^{\frac{1}{2}} \\
 &= (7,40 + 0,19 + 7,81)^{\frac{1}{2}} \\
 &= 3,92
 \end{aligned}$$

Dengan cara yang sama dilakukan perhitungan jarak antar C1 dengan provinsi lainnya, hasil perhitungannya dapat dilihat pada Tabel 2.4.

Tabel 2. 4 Contoh hasil perhitungan jarak *Euclidean* C1

No	Provinsi	C1
1	DKI Jakarta	3,92
2	Maluku	4,50
3	Papua	4,98

Maka menghasilkan matriks jarak *Euclidean* baru yang berukuran 4×4 sebagai berikut.

	DKI Jakarta	Maluku	Papua	C1
DKI Jakarta	0,00	7,02	7,80	3,92
Maluku	7,02	0,00	3,24	4,50
Papua	7,80	3,24	0,00	4,98
C1	3,92	4,50	4,98	0,00

Perbedaan antara matriks jarak ini dengan matriks jarak analognya (Tabel 2.3) hanya terdapat pada jarak antara C1 dan provinsi lainnya. Pada matrik jarak ini Provinsi Maluku dan Papua memiliki jarak terkecil yaitu 3,24, sehingga C2 adalah kelompok kedua yang beranggotakan Maluku dan Papua. Perhitungan *centroid* variabel X_1 untuk C2 adalah sebagai berikut.

$$\bar{X}_{1(4,5)} = \frac{(1 \times 66,45) + (1 \times 66,23)}{1 + 1} = 66,34$$

Dengan cara yang sama lakukan perhitungan *centroid* C2 untuk variabel lainnya. Maka didapatkan nilai *centroid* C2 yang dapat dilihat pada Tabel 2.4 berikut ini.

Tabel 2. 5 Contoh hasil perhitungan nilai *centroid* C2

No	Provinsi	$\bar{X}_1$	$\bar{X}_2$	$\bar{X}_3$
4 & 5	C2	66,34	12,57	9,44

Selanjutnya jarak kembali dihitung berdasarkan data provinsi DKI Jakarta, C1, dan C2. Perhitungan jarak Provinsi DKI Jakarta dengan C2 adalah sebagai berikut.

$$\begin{aligned}
 d_{(2,C2)} &= ((73,32 - 66,34)^2 + (13,08 - 12,57)^2 + (11,31 - 9,44)^2)^{\frac{1}{2}} \\
 &= (6,98 + 0,51 + 1,87)^{\frac{1}{2}} \\
 &= 7,24
 \end{aligned}$$

Dengan cara yang sama dilakukan perhitungan jarak antar C2 dengan C1, sehingga hasil perhitungannya dapat dilihat pada Tabel 2.6.

Tabel 2. 6 Contoh hasil perhitungan jarak Euclidean C2

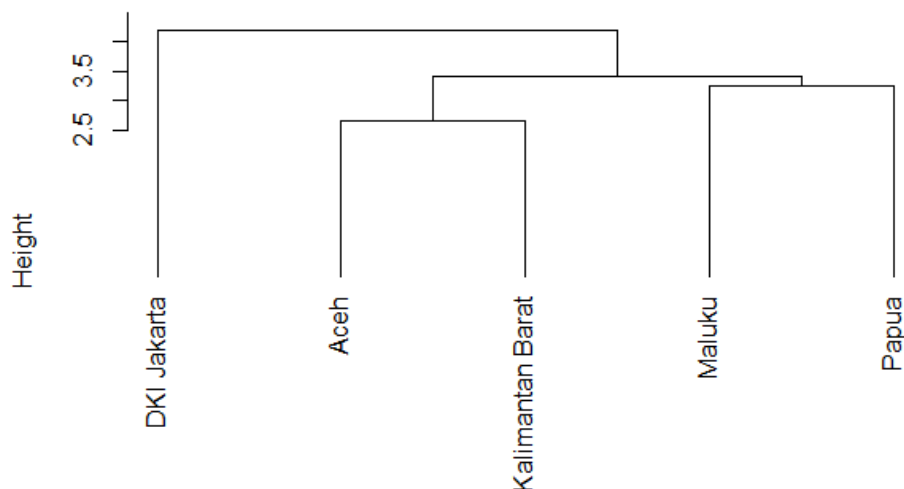
No	Provinsi	C2
1	DKI Jakarta	7,24
2	C1	3,41

Perhitungan yang sama dilakukan sampai semua provinsi berada pada satu kelompok. Tahapan pengelompokan Provinsi Aceh, DKI Jakarta, Kalimantan Barat, Maluku dan Papua ditunjukkan pada Tabel 2.5.

Tabel 2. 7 Tahapan-tahapan pengelompokan untuk contoh metode centroid

Tahap	Kelompok yang terbentuk	Jarak	Jumlah Kelompok
1	C1 = Aceh, Kalimantan Barat	2,66	4
2	C2 = Maluku, Papua	3,24	3
3	C3 = C1 (Aceh, Kalimantan Barat), C2 (Maluku Papua)	3,41	2
4	C4 = C3 (Aceh, Kalimantan Barat, Maluku, Papua), DKI Jakarta	4,18	1

Seluruh proses dalam contoh pengelompokan dengan metode *centroid* membentuk dendrogram yang ditunjukkan pada Gambar 2.2.



Gambar 2. 2. Contoh dendrogram untuk metode *centroid*

Berdasarkan Gambar 2.2, jika dendrogram dipotong berdasarkan jarak terjauh (4,18) akan dihasilkan sebanyak 2 kelompok. Kelompok 1 beranggotakan DKI Jakarta dan kelompok 2 terdiri dari Aceh, Kalimantan Barat, Maluku, dan Papua. Jika dendrogram dipotong pada jarak 3,41, maka kelompok yang terbentuk adalah 3 kelompok. Kelompok 1 terdiri dari Aceh dan Kalimantan Barat, kelompok 2 terdiri dari Maluku dan Papua, dan kelompok 3 adalah DKI Jakarta. Untuk menentukan jumlah kelompok yang tepat dapat melakukan validasi jumlah kelompok.

2.5 Indeks Pembangunan Manusia

IPM menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya. IPM merupakan indikator penting untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia di suatu wilayah, (BPS, 2023a). Pengukuran IPM di Indonesia mengacu pada tiga dimensi yang diperkenalkan oleh *United Nations Development Programme* (UNDP) pada tahun 1990. Ketiga dimensi itu terdiri dari umur panjang dan hidup sehat, pengetahuan dan standar hidup yang layak (BPS, 2023).

Sejak tahun 2014 angka IPM di Indonesia disajikan periodik setiap tahun pada tingkat nasional, provinsi, dan kabupaten/kota. Penyajian IPM memungkinkan setiap daerah untuk mengetahui peta pembangunan manusia di daerahnya, baik capaian, posisi, kecepatan antar daerah (BPS, 2023).

Menurut BPS, capaian pembangunan manusia di suatu wilayah pada waktu tertentu, dapat dikelompokkan ke dalam empat kelompok menurut status capaiannya (BPS, 2023). Adapun pengelompokkan tersebut adalah:

- a) Rendah, jika $IPM < 60$
- b) Sedang, jika $60 \leq IPM < 70$
- c) Tinggi, jika $70 \leq IPM < 80$
- d) Sangat tinggi, jika $IPM \geq 80$

(BPS, 2023)

Nilai IPM berkisar antara 0 – 100, apabila nilai IPM di suatu wilayah semakin mendekati angka 100 maka semakin baik pembangunan manusia di wilayah tersebut (BPS, 2023a). Perhitungan IPM dilakukan dengan menghitung indeks dari indikator

pembentuk IPM. Berdasarkan nilai indeks pengetahuan, kesehatan dan pengeluaran, didapatkan rumus perhitungan IPM sebagai berikut.

$$IPM = \left(\sqrt[3]{I_{pengetahuan} \times I_{kesehatan} \times I_{pengeluaran}} \right) \times 100$$

Keterangan:

IPM : Indeks Pembangunan Manusia

$I_{pengetahuan}$: Indeks pengetahuan

$I_{kesehatan}$: Indeks kesehatan

$I_{pengeluaran}$: Indeks pengeluaran

Setiap indikator pembentuk IPM dapat dimanfaatkan untuk mengukur keberhasilan pembangunan kualitas hidup manusia (BPS, 2022).

2.5.1 Indikator Pembentuk IPM

Terdapat empat indikator penyusun IPM yaitu HLS, RLS, AHH, dan PPP. Keempat indikator yang akan digunakan tersebut akan dijelaskan sebagai berikut.

a. Harapan Lama Sekolah (HLS)

HLS adalah harapan jumlah tahun lamanya sekolah yang akan dijalani oleh anak usia 7 tahun di masa mendatang. Menurut BPS, HLS yang baik adalah yang mendekati atau lebih tinggi dari 12 tahun yang setara dengan menyelesaikan pendidikan menengah atas/ sederajat (BPS, 2023a). HLS digunakan untuk mengetahui kondisi pembangunan sistem pendidikan di berbagai jenjang yang ditunjukkan dalam bentuk jumlah tahun lamanya pendidikan, yang diharapkan dapat dicapai oleh setiap anak (BPS, 2023).

$$HLS_a^t = FK \times \sum_{i=a}^n \frac{E_i^t}{P_i^t}$$

Keterangan:

HLS_a^t : Harapan lama sekolah pada usia a di tahun t

E_i^t : Jumlah penduduk usia i yang bersekolah pada tahun t

P_i^t : Jumlah penduduk usia i pada tahun t

FK : Faktor koreksi persantren

i : usia ($a, a + 1, \dots, n$)

b. Rata-rata Lama Sekolah (RLS)

RLS adalah rata-rata lamanya waktu yang digunakan oleh penduduk dalam menjalani pendidikan formal. menggambarkan tingkat pendidikan stok modal manusia yang dimiliki oleh suatu wilayah. Cakupan penduduk untuk menghitung RLS adalah penduduk berumur 25 tahun ke atas, dengan asumsi bahwa proses pendidikan seseorang sudah berakhir sebelum mencapai umur 25 tahun (BPS, 2023a). Menurut BPS, RLS yang baik adalah yang mendekati atau lebih tinggi dari 12 tahun yang setara dengan menyelesaikan pendidikan menengah atas/ sederajat (BPS, 2023).

$$RLS = \frac{1}{v} \sum_{i=1}^n R_i$$

Keterangan:

RLS : Rata-rata lama sekolah penduduk usia 25 tahun ke atas

R_i : Lama sekolah penduduk i yang berusia 25 tahun

v : Jumlah penduduk usia 25 tahun ke atas

c. Angka Harapan Hidup (AHH)

AHH merupakan indikator yang merefleksikan dimensi umur panjang dan hidup sehat. Penggunaan indikator ini didasarkan pada kenyataan bahwa umur panjang merupakan sesuatu yang tak ternilai dan dapat terwujud jikalau manusia mendapatkan nutrisi yang cukup dan kesehatan yang baik. Dilihat dari sisi definisi, AHH merupakan perkiraan rata-rata lamanya waktu yang dapat dijalani oleh seseorang selama hidupnya dalam tahun (BPS, 2023). AHH yang baik adalah yang berada di atas 70 tahun (BPS, 2023a).

d. Pengeluaran per kapita disesuaikan (PPP)

Dalam mengacu pada standar UNDP, Indonesia melakukan penyesuaian pada indikator Pendapatan Nasional Bruto (PNB) per kapita yang merupakan cerminan dimensi standar hidup yang layak karena tidak tersedia di tingkat provinsi dan kabupaten/kota, maka digantikan dengan indikator PPP. Pergantian indikator ini berdampak positif bagi hasil penghitungan IPM dimana pendapatan dan tingkat kesejahteraan masyarakat dapat tercermin melalui indikator PPP. Indikator PPP menggambarkan kemampuan ekonomi masyarakat dalam

memenuhi kebutuhan pokok seperti pangan, sandang, papan, pendidikan, layanan kesehatan, dan keperluan lainnya yang dapat meningkatkan kesejahteraan. Menurut BPS, PPP yang baik di Indonesia pada tahun 2022 adalah lebih dari atau sama dengan Rp 11.000.000,00 (BPS, 2023a). Pengeluaran per kapita yang “baik” untuk wilayah Papua dapat bervariasi tergantung pada konteks ekonomi, sosial, dan lingkungan. Untuk menentukan angka yang spesifik dianggap baik, dapat memperhatikan garis kemiskinan. Angka pengeluaran yang baik akan jauh di atas garis kemiskinan (BPS, 2023b). Pengeluaran riil per kapita yang telah disesuaikan dihitung menggunakan data Survei Sosial Ekonomi Nasional (SUSENAS) modul konsumsi bulan Maret, indeks harga konsumen, serta harga komoditas non-makanan hasil survei harga konsumen. (BPS, 2023a).

$$PDB_j = \prod_{i=1}^m \left(\frac{P_{ij}}{P_{ik}} \right)^{\frac{1}{m}}$$

Keterangan:

PDB_j : Paritas daya beli di wilayah j

P_{ij} : Harga komoditas i di kabupaten/kota j

P_{ik} : Harga komoditas i di Papua

m : Jumlah komoditas

2.5.2 Penelitian Terdahulu

Penelitian terdahulu dicantumkan dengan maksud sebagai pendukung dalam penelitian yang akan dilakukan oleh peneliti, dimana penelitian-penelitian ini berkaitan dengan indikator pembentuk IPM.

1. Penelitian yang dilakukan oleh Sunjadeva, R. (2020) tentang Pengelompokan Kabupaten/kota di Provinsi Papua Berdasarkan Indikator IPM. Diketahui bahwa variabel yang digunakan oleh peneliti yaitu HLS, RLS, AHH, dan PPP. Berdasarkan penelitian ini, didapatkan hasil pengelompokan yaitu 5 kelompok. Kelompok 1 beranggotakan 5 kabupaten yaitu Merauke, Biak Numfor, Jayapura, Mimika, dan Nabire. Kelompok 2 beranggotakan 10 kabupaten yaitu Paniai, Asmat, Supiori, Intan Jaya, Puncak Jaya, Pegunungan Bintang, Puncak, Mappi, Sarmi, dan Dogiyai. Kelompok 3 beranggotakan 1 kota yaitu Kota

Jayapura. Kelompok 4 beranggotakan 8 kabupaten yaitu Yahukimo, Lanny Jaya, Deiyai, Tolikara, Mamberamo Raya, Mamberamo Tengah, Nduga, dan Yalimo. Kelompok 5 beranggotakan 5 kabupaten yaitu Jayawijaya, Waropen, Kepulauan Yapen, Boven Digoel, dan Keerom. Berdasarkan karakteristik masing-masing kelompok yang terbentuk, diketahui bahwa kelompok 3 merupakan kelompok dengan nilai indikator IPM yang tertinggi, dan kelompok 4 merupakan kelompok dengan nilai indikator IPM terendah untuk semua indikator IPM.

2. Penelitian yang dilakukan oleh Qori'atunnadyah (2023) tentang Pengelompokan Kabupaten/kota di Provinsi Jawa Timur Berdasarkan Indikator IPM. Penelitian ini menggunakan 4 variabel yaitu RLS, HLS, AHH, dan PPP. Berdasarkan penelitian ini, didapatkan hasil pengelompokan yaitu 4 kelompok. Kelompok 1 terdiri 9 kabupaten, kelompok 2 terdiri dari 13 kabupaten, kelompok 3 terdiri dari 9 kabupaten, dan kelompok 4 terdiri dari 7 kabupaten/kota. Berdasarkan analisis karakteristik dari masing-masing kelompok, diketahui bahwa kelompok 4 memiliki nilai indikator IPM tertinggi. Sedangkan kelompok dengan nilai indikator IPM terendah adalah kelompok 1.

Berdasarkan penelitian-penelitian terdahulu tentang IPM dan referensi yang didapatkan dari BPS Indonesia dan BPS Provinsi Papua, penelitian ini akan menggunakan 4 variabel yaitu RLS, HLS, AHH, dan PPP. Variabel-variabel tersebut adalah indikator berdasarkan tiga dimensi pembentuk IPM.