

## Analisis Kesenjangan Regional di Sektor Pertanian Indonesia Tahun 2023 Melalui Pendekatan *Cluster*

Melynda Isaura<sup>1\*</sup>, Fauzi Dwi Aryasatyawan<sup>2</sup>, Sri Pingit Wulandari<sup>3</sup>

<sup>1-3</sup>Departemen Statistika Bisnis, Fakultas Vokasi,  
Institut Teknologi Sepuluh Nopember, Indonesia

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

Korespondensi Penulis: [melyndaisauraaa@gmail.com](mailto:melyndaisauraaa@gmail.com)\*

**Abstract.** *Indonesia is an agricultural country with great potential in the agricultural sector, both as a source of livelihood and as an economic driver. This sector has an important function in food security, improving farmers' welfare, and preserving the environment. Given Indonesia's regional diversity, it is important to develop policies that suit the agricultural characteristics of each region. This research uses the clustering method, a technique that groups data based on similar characteristics so that objects in one group (cluster) have high similarities, while between clusters they are different. This research applies the clustering method to group provinces in Indonesia based on agricultural sector factors in 2023, using hierarchical and non-hierarchical techniques. The results show that the optimum clusters obtained using both single linkage and K-Means formed five optimal clusters. The Java Island cluster shows high productivity, while Riau, East Kalimantan and North Kalimantan have quite high productivity with minimal labor. The largest clusters have a lot of unmanaged land, increasing employment opportunities. Significant variables influence grouping.*

**Keywords:** *Clustering, Hierarchy, Non-Hierarchy, Agricultural Sector.*

**Abstrak.** Indonesia adalah negara agraris dengan potensi besar di sektor pertanian, baik sebagai sumber penghidupan maupun sebagai penggerak ekonomi. Sektor ini memiliki fungsi penting dalam ketahanan pangan, peningkatan kesejahteraan petani, dan pelestarian lingkungan. Mengingat keragaman wilayah Indonesia, penting untuk menyusun kebijakan yang sesuai dengan karakteristik pertanian di setiap daerah. Penelitian ini menggunakan metode clustering, teknik yang mengelompokkan data berdasarkan kemiripan karakteristik sehingga objek dalam satu kelompok (cluster) memiliki kemiripan tinggi, sedangkan antar cluster berbeda. Penelitian ini menerapkan metode clustering untuk mengelompokkan provinsi di Indonesia berdasarkan faktor-faktor sektor pertanian pada tahun 2023, menggunakan teknik hirarki dan non-hirarki. Hasilnya menunjukkan bahwa klaster optimum yang didapatkan baik dengan single linkage maupun K-Means terbentuk lima klaster optimal. Klaster Pulau Jawa menunjukkan produktivitas tinggi, sementara Riau, Kalimantan Timur, dan Kalimantan Utara memiliki produktivitas cukup tinggi dengan tenaga kerja minim. Klaster terbesar memiliki banyak lahan tak terkelola, meningkatkan peluang kerja. Variabel signifikan memengaruhi pengelompokkan.

**Kata Kunci:** *Clustering, Hirarki, Non-Hirarki, Sektor Pertanian.*

### 1. PENDAHULUAN

Indonesia dikenal sebagai negara agraris dengan lahan yang sangat luas, yang dapat dimanfaatkan masyarakat sekitar sebagai sumber penghidupan. Namun, sektor agraris atau pertanian di Indonesia tidak hanya berfungsi sebagai mata pencaharian, tetapi juga berpotensi meningkatkan perekonomian negara (S. I. Kusumaningrum, 2019). Komoditas pertanian Indonesia memiliki daya saing yang cukup tinggi di pasar internasional. Sektor pertanian memiliki nilai multifungsi yang besar dalam memperkuat ketahanan pangan, meningkatkan kesejahteraan petani, dan menjaga kelestarian lingkungan. Fungsi-fungsi ini mencakup aspek produksi atau ketahanan pangan, peningkatan kesejahteraan petani dan pengurangan

kemiskinan, serta pelestarian lingkungan. Bagi Indonesia, peran penting sektor pertanian ini perlu diperhatikan dalam penyusunan kebijakan terkait struktur insentif di sektor pertanian (K. D. Wahyudi, 2012). Namun, pada dasarnya, Indonesia adalah negara kepulauan terbesar di dunia, memiliki keberagaman yang signifikan di berbagai bidang, termasuk dalam sektor pertanian. Oleh karena itu, untuk menyusun kebijakan yang sesuai diperlukan identifikasi lebih mendalam terkait karakteristik sektor pertanian antar wilayah.

*Clustering* adalah teknik untuk mengelompokkan data. *Clustering* membagi sekumpulan objek data menjadi kelompok yang disebut *cluster*, di mana data dalam satu *cluster* memiliki kemiripan yang tinggi, sementara kemiripan antar *cluster* sangat rendah. Objek dalam satu kluster memiliki karakteristik serupa, berbeda dari objek di kluster lainnya. Jarak sering digunakan untuk menentukan kesamaan atau perbedaan antara dua vektor fitur berdasarkan ranking, semakin kecil nilai ranking, semakin tinggi kesamaannya. *Clustering* umumnya dibagi menjadi dua jenis yakni kluster hirarki yang mengelompokkan data dalam struktur hirarki, dengan menggabungkan dua objek terdekat di setiap iterasi. Sementara itu, kluster non-hirarki membagi data menjadi beberapa kluster tanpa struktur hirarki, dengan setiap kluster memiliki titik pusat (*centroid*). Tujuan utama kluster non-hirarki adalah meminimalkan jarak antara data dan pusat kluster (E. Irwansyah and M. Faisal, 2015).

Pada penelitian ini akan dilakukan *clustering* pada provinsi di Indonesia berdasarkan faktor-faktor sektor pertanian tahun 2023. *Clustering* menggunakan metode yang paling optimum baik secara hirarki maupun non-hirarki. *Clustering* pada penelitian ini bertujuan untuk membagi provinsi-provinsi di Indonesia menjadi beberapa kluster sehingga akan diketahui karakteristik sektor pertanian pada setiap kluster yang terbentuk. Hasil analisis diharapkan dapat memberikan pemahaman mendalam mengenai perbedaan karakteristik sektor pertanian di Indonesia sehingga kemudian dapat bermanfaat sebagai bahan pertimbangan dalam menyusun kebijakan strategis untuk meningkatkan produktivitas pertanian.

## 2. TINJAUAN PUSTAKA

### Statistika Deskriptif

Statistika deskriptif merupakan ilmu statistika yang membahas cara-cara mengumpulkan data, bagaimana melakukan penyederhanaan terhadap data baik yang berupa angka maupun yang bukan, serta bagaimana meringkas dan menyajikannya dalam bentuk tabel dan grafik untuk memperoleh informasi yang lebih menarik, lebih berguna, dan lebih mudah dipahami oleh pengguna (S. Siregar, 2017).

### 1) *Mean*

Nilai *mean* atau rata-rata merupakan suatu nilai yang dianggap dapat mewakili suatu kumpulan data. Nilai *mean* hanya dapat memberikan gambaran kasar mengenai di sekitar mana nilai-nilai pengamatan lainnya tersebar. Nilai *mean* sangat mudah dipengaruhi oleh keberadaan nilai ekstrem, yaitu nilai yang sangat besar atau sangat kecil (S. Siregar, 2017). Untuk memperoleh nilai *mean*, dapat dihitung dengan menggunakan rumus berikut.

$$\bar{X} = \frac{\sum X_i}{N} \quad (1)$$

Keterangan:

$\bar{X}$  = *Mean*

$X_i$  = Nilai data ke-i

$N$  = Banyaknya data

### 2) *Standar Deviasi*

Simpangan baku (standar deviasi) adalah nilai yang menunjukkan tingkat variasi kelompok data atau ukuran standar penyimpangan dari nilai rata-ratanya. Lambang standar deviasi untuk populasi =  $\sigma$  (tho) dan untuk sampel =  $s$  (E. Hendrawaty, 2006). Standar deviasi merupakan besarnya penyebaran variabel random diantara rata-ratanya. Semakin besar penyebarannya, semakin besar standar deviasi tersebut (I. Ghozali, 2018). Rumus diperoleh menggunakan persamaan berikut.

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (2)$$

Keterangan:

$S$  = Standar deviasi

$\bar{X}$  = Nilai rata-rata

$N$  = Jumlah data

### 3) *Minimum*

Minimum adalah nilai terendah dari suatu data. Nilai minimum didapatkan dengan mudah apabila kita mengurutkan data, data yang paling kecil itulah yang disebut nilai minimum (S. Siregar, 2017).

### 4) *Maksimum*

Maksimum adalah nilai tertinggi dari suatu data. Nilai maksimum didapatkan dengan mudah apabila kita mengurutkan data, data yang paling besar itulah yang disebut nilai maksimum (S. Siregar, 2017).

## Asumsi Analisis Diskriminan

Asumsi-asumsi dalam analisis diskriminan yang harus dipenuhi meliputi asumsi uji distribusi normal multivariat dan uji independensi. Kedua asumsi tersebut dijelaskan sebagai berikut.

### 1) Pengujian Asumsi Distribusi Normal Multivariat

Uji distribusi normal multivariat dilakukan untuk melihat distribusi data memenuhi asumsi distribusi normal multivariat atau tidak. Uji asumsi distribusi normal multivariat dalam penelitian ini khususnya akan dilakukan pengujian t proporsi dimana data dikatakan berdistribusi normal multivariat jika memiliki nilai proporsi 45% – 55%. Nilai t proporsi sendiri didapatkan dari nilai  $d_j^2$  dengan  $q_{c,p}$  (Sudaryono, 2021). Perumusan hipotesis uji ini adalah sebagai berikut.

$$H_0: F_n(X) = F_0(X) \text{ (Data berdistribusi normal multivariat)}$$

$$H_1: F_n(X) \neq F_0(X) \text{ (Data tidak berdistribusi normal multivariat)}$$

Daerah penolakan: Tolak  $H_0$  jika T-proporsi < 0,45 atau T-proporsi > 0,55

Statistik Uji:

$$d_j^2 = (x_j - \bar{x})S^{-1}(x_j - \bar{x}) \quad (3)$$

Keterangan:

$d_j^2$  = Nilai kuadrat ke- $j$

$x_j$  = Objek pengamatan ke- $j$

$\bar{x}$  = Rata-rata nilai variabel  $x$

$S^{-1}$  = Invers matriks varians kovarians

### 2) Pengujian Asumsi Independensi

Uji *Bartlett* adalah salah satu metode yang digunakan untuk mengevaluasi homogenitas atau keseragaman varians galat dalam analisis statistik. Keberadaan asumsi kehomogenan variansi galat menjadi krusial karena ketidakseragaman varians galat bisa mengakibatkan hasil yang tidak akurat dalam pengujian MANOVA. Oleh karena itu, sebelum melakukan analisis MANOVA, sangat penting untuk memeriksa apakah asumsi ini terpenuhi atau tidak (Sudaryono, 2021).

Hipotesis:

$H_0$  : Matriks korelasi identik dengan matriks identitas antar variabel independen.

$H_1$  : Matriks korelasi tidak identik dengan matriks identitas antar variabel dependen.

Taraf Signifikan:  $\alpha$

Daerah Penolakan:  $H_0$  ditolak apabila  $X^2_{hitung} > X^2_{\alpha,df}$  atau P-value  $< \alpha$

Statistik Uji :

$$X^2 = (ln.n)\{B - \sum dk \log si^2\} \quad (4)$$

Keterangan:

$n$  = jumlah data

$$B = \sum(dk) \log s^2; \text{ yang mana } s^2 = \frac{\sum(dk si^2)}{\sum dk}$$

$si^2$  = varians data untuk setiap kelompok ke  $- i$

$dk$  = derajat bebas

### Analisis Cluster

Analisis kelompok atau yang biasa dikenal sebagai *cluster analysis* adalah salah satu teknik statistik yang bertujuan untuk mengelompokkan objek kedalam suatu kelompok. Objek yang berada dalam satu kelompok akan memiliki kesamaan yang tinggi dibandingkan dengan objek yang berada di kelompok lain (R. A. Johnson and D. Wichern, 2014). Ada dua metode pengelompokan (*cluster analysis*) yaitu *hierarchical clustering* dan *partitioning* atau biasa disebut *non-hierarchical clustering*.

#### 1) Analisis Cluster Hierarki

Analisis *cluster* hierarki adalah suatu metode yang membentuk tingkatan tertentu seperti pada struktur pohon karena proses pengklasteran dilakukan secara bertahap atau bertingkat. Hasil pengklasteran dengan metode hierarki dapat disajikan dalam bentuk *dendrogram*. *Dendrogram* adalah representasi visual dari langkah-langkah dalam analisis *cluster* terbentuk dan nilai koefisien jarak pada setiap langkah (B. Simamora, 2005). Metode hierarki yang digunakan dalam penelitian ini adalah metode *average linkage*. Pada metode *average linkage*, jarak antara dua *cluster* dianggap sebagai jarak rata-rata antara semua anggota dalam satu *cluster* dengan semua anggota *cluster* lain [11]. Metode *average linkage* diperoleh dengan menggunakan persamaan berikut.

$$d_{(ij)k} = \frac{\sum a \sum b d_{ab}}{N_{ij}N_k} \quad (5)$$

Keterangan:

$d_{ab}$  = Jarak antara objek i pada cluster (IJ) dan obyek b pada  
cluster K

$N_{ij}$  = Jumlah item pada cluster (IJ)

$N_k$  = Jumlah item pada cluster (IJ) dan K

## 2) Analisis Cluster Non-Hierarki

Analisis *cluster* dengan metode non-hierarki merupakan metode *cluster* yang menentukan jumlah *cluster* secara manual (A. Baroroh, 2012). Teknik analisis *cluster* non-hierarki dirancang untuk mengelompokkan item, bukan variabel, menjadi kumpulan *K cluster*. Banyaknya *cluster* (*k*), ditentukan terlebih dahulu untuk memulai prosedur pengklasteran. Metode analisis *cluster* non-hierarki yang digunakan dalam penelitian ini adalah *K-Means* (N. Amah, dkk, 2017). Metode *K-Means* merupakan metode untuk mengklasifikasikan atau mengelompokkan objek-objek (data) ke dalam *K*-grup (*Cluster*) berdasarkan atribut tertentu. Pengelompokan data dilakukan dengan memperhitungkan jarak terdekat antara data-data (objek observasi) dengan pusat *cluster* (*centroid*). Prinsip utama dari metode ini adalah menyusun rata-rata (*mean*) dari sekumpulan data berdimensi *N*, dimana metode ini mensyaratkan nilai *K* sudah diketahui sebelumnya (*apriori*) (D. R. Ningrat, dkk, 2016). Proses analisis klaster menggunakan metode *K-Means* adalah sebagai berikut.

- a. Menentukan *k* sebagai jumlah klaster yang ingin dibentuk.
- b. Mengalokasikan data ke dalam klaster secara acak.
- c. Menentukan pusat klaster (*centroid*) dari data yang ada pada masing-masing klaster dengan persamaan sebagai berikut.

$$C_{kj} = \frac{X_{1j} + X_{2j} + \dots + X_{nj}}{n} \quad (6)$$

- d. Menentukan jarak setiap objek dengan setiap *centroid* dengan perhitungan jarak setiap objek dengan setiap *centroid* menggunakan jarak *euclidean*. Rumus jarak *euclidean* adalah sebagai berikut.

$$d(X_i, X_g) = \sqrt{\sum_{j=1}^p (X_{ij} - X_{gj})^2} \quad (7)$$

- e. Mengalokasikan masing-masing data ke rata-rata terdekat.
- f. Mengulangi kembali langkah 3-5 sampai tidak ada lagi perpindahan objek.

## 3) Penentuan Cluster Optimum

Jumlah kelompok optimum dapat diketahui dengan kriteria nilai *Pseudo F-statistics*. Berikut adalah persamaan yang digunakan untuk menghitung nilai *Pseudo F- statistics*.

$$Pseudo - F = \frac{\left(\frac{R^2}{k-1}\right)}{\left(\frac{1-R^2}{n-k}\right)} \quad (8)$$

dengan,

$$R^2 = \frac{(SST-SSW)}{SST} \quad (9)$$

$$SST = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (X_{ijk} - \bar{X}_j)^2 \quad (10)$$

$$SSW = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (X_{ijk} - \bar{X}_{jk})^2 \quad (11)$$

Keterangan:

$SST$  = Total jumlah dari kuadrat jarak sampel terhadap rata-rata keseluruhan

$SSW$  = Total jumlah dari kuadrat jarak sampel terhadap rata-rata kelompoknya

$n$  = Banyaknya sampel

$c$  = Banyaknya variabel

$p$  = Banyaknya kelompok

$X_{ijk}$  = Sampel ke- $i$  pada variabel ke- $j$  kelompok ke- $k$

$\bar{X}_j$  = Rata-rata seluruh sampel pada variabel ke- $j$

$\bar{X}_{jk}$  = Rata-rata sampel pada variabel ke- $j$  dan kelompok ke- $k$

Nilai *Pseudo F-statistics* tertinggi menunjukkan bahwa jumlah kelompok yang digunakan untuk membagi data secara optimal, dimana keragaman dalam kelompok sangat homogen sedangkan antar kelompok sangat heterogen. Salah satu metode yang digunakan untuk memilih metode *cluster* terbaik adalah menghitung rata-rata persebaran *Internal Cluster Dispersion Rate (icdrate)* terhadap partisi keseluruhan. *Icdrate (internal cluster dispersion)* menggambarkan tingkat dispersi atau perbedaaan dalam *cluster*. Formula untuk menghitung *icdrate* terdapat di Persamaan 8. Semakin kecil nilai *icdrate* maka semakin baik hasil pengelompokannya. (I. N. L. Fitriana, 2021).

$$Icdrate = 1 - \frac{SSB}{SST} = 1 - \frac{SST-SSW}{SST} = 1 - R^2 \quad (12)$$

Keterangan:

$SSB$  = *Sum Square Between*

$R^2$  = *Recovery Rate*

#### a. *Produktivitas Padi*

Produktivitas padi adalah ukuran efisiensi yang menggambarkan seberapa banyak hasil padi yang dapat diproduksi per unit luas lahan dalam waktu tertentu, biasanya dinyatakan dalam ton per hektar. Padi merupakan salah satu tanaman pangan utama di banyak negara, terutama di Asia, dan merupakan sumber makanan bagi sebagian besar populasi dunia. Faktor-faktor yang memengaruhi produktivitas padi meliputi pemilihan varietas, teknik budidaya, manajemen irigasi, penggunaan pupuk, serta pengendalian hama dan penyakit (S. Snapp, 2017).

*b. Produksi Padi*

Produksi padi merupakan proses budidaya tanaman padi yang meliputi pemilihan varietas, persiapan lahan, penanaman, pemeliharaan, dan panen. Faktor-faktor yang mempengaruhi produksi padi termasuk iklim, ketersediaan sumber daya alam, dan teknologi pertanian. Pertanian yang baik, seperti penggunaan varietas unggul dan sistem irigasi yang efisien, dapat meningkatkan hasil dan kualitas padi (S. Snapp, 2017).

*c. Nilai Tambah Pertanian*

Nilai tambah pertanian adalah ukuran yang menunjukkan kontribusi ekonomi dari setiap pekerja dalam sektor pertanian, dihitung sebagai selisih antara nilai output dan input dalam proses produksi. Ukuran ini penting untuk menilai produktivitas, merancang kebijakan yang meningkatkan efisiensi, dan mengidentifikasi kebutuhan pelatihan bagi tenaga kerja. Beberapa faktor yang mempengaruhi nilai tambah ini termasuk penggunaan teknologi modern, pendidikan dan keterampilan tenaga kerja, serta akses ke pasar (Surya, 2018).

*d. Tenaga Kerja Informal*

Tenaga kerja informal adalah pekerja yang terlibat dalam kegiatan pertanian tanpa perjanjian formal atau perlindungan hukum, seperti buruh harian atau petani kecil. Tenaga kerja informal tidak terdaftar dalam sistem ketenagakerjaan, bekerja dengan jam yang tidak tetap, dan menghadapi kondisi kerja yang buruk tanpa akses ke perlindungan hukum. Meskipun menerima upah rendah dan bekerja dalam situasi tidak menentu tenaga kerja informal berkontribusi signifikan terhadap produktivitas sektor pertanian dan ketahanan pangan, terutama di negara berkembang Soetomo, 2019).

*e. Lapangan Kerja Informal*

Lapangan kerja informal dalam pertanian mencakup pekerjaan tanpa kontrak formal atau perlindungan hukum, seperti buruh tani, petani musiman, dan pekerja lepas dalam pengolahan dan pemasaran produk pertanian. Pekerja informal ini sering kali tidak terdaftar dalam statistik ketenagakerjaan, memiliki jam kerja yang fleksibel dan tidak memiliki akses ke perlindungan sosial seperti asuransi kesehatan atau pensiun, serta sering menerima upah rendah dan tidak menentu. Meskipun demikian memiliki peran penting dalam ketahanan pangan dan ekonomi lokal terutama di daerah pedesaan serta membantu mengurangi kemiskinan (Rizki, 2020).

### 3. METODOLOGI PENELITIAN

#### Sumber Data

Data yang digunakan dalam praktikum ini merupakan data sekunder diperoleh dari *website* Badan Pusat Statistika (BPS). Data merupakan faktor-faktor yang memengaruhi sektor pertanian di 34 Provinsi di Indonesia pada tahun 2023. Data diakses oleh peneliti pada hari Selasa, 4 November 2024 di Kota Surabaya, Jawa Timur.

#### Variabel Penelitian

Variabel penelitian yang digunakan dalam praktikum ini ditunjukkan pada Tabel 1 sebagai berikut.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Skala Pengukuran	Satuan
X <sub>1</sub>	Produktivitas Padi	Rasio	Kuintal/Hektar
X <sub>2</sub>	Produksi Padi	Rasio	Ton
X <sub>3</sub>	Nilai Tambah Pertanian per Tenaga Kerja	Rasio	Rupiah
X <sub>4</sub>	Proporsi Lapangan Kerja Informal Sektor Pertanian	Rasio	Persen

#### Struktur Data

Struktur data yang digunakan pada praktikum ini dijelaskan pada Tabel 2 sebagai berikut.

Tabel 2. Struktur Data

Provinsi	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>
1	X <sub>1,1</sub>	X <sub>1,2</sub>	X <sub>1,3</sub>	X <sub>1,4</sub>
2	X <sub>2,1</sub>	X <sub>2,2</sub>	X <sub>2,3</sub>	X <sub>2,4</sub>
⋮	⋮	⋮	⋮	⋮
34	X <sub>34,1</sub>	X <sub>34,2</sub>	X <sub>34,3</sub>	X <sub>34,4</sub>

#### Langkah Analisis

Langkah analisis dalam praktikum ini adalah sebagai berikut.

1. Mengumpulkan data sekunder faktor-faktor yang memengaruhi sektor pertanian melalui *website bps.go.id*.
2. Mendeskripsikan karakteristik data faktor-faktor yang memengaruhi sektor pertanian menggunakan tabel statistika deskriptif.
3. Melakukan pengujian asumsi analisis kluster yang terdiri dari uji distribusi normal multivariat dan uji independensi.
4. Melakukan standarisasi pada data faktor-faktor yang memengaruhi sektor pertanian.
5. Mengidentifikasi metode dan jumlah kluster terpilih berdasarkan nilai *Pseudo-F* dan *Icdrate* dengan langkah sebagai berikut.
  - a. Menentukan metode analisis kluster hierarki dan jumlah kluster optimum
  - b. Menentukan jumlah kluster optimum untuk analisis kluster non-hierarki (*K-Means*).

1. Menginterpretasikan hasil analisis kluster hierarki dan kluster non-hierarki (*K-Means*).
2. Menarik kesimpulan dan saran.

#### 4. ANALISIS DAN PEMBAHASAN

Bab ini akan membahas mengenai karakteristik data, pengujian asumsi analisis *cluster*, serta hasil pengelompokkan secara hierarki dan non-hierarki pada faktor-faktor sektor pertanian di Indonesia tahun 2023.

##### Karakteristik Faktor-Faktor Sektor Pertanian

Karakteristik data dari faktor-faktor sektor pertanian di Indonesia tahun 2023 yang dianalisis menggunakan statistika deskriptif untuk mengetahui ukuran sampel yang diteliti seperti *mean*, standar deviasi, nilai maksimum, dan nilai minimum dari masing-masing variabel yang hasilnya dijelaskan pada Tabel 3 sebagai berikut.

Tabel 3. Karakteristik Data

Variabel	Mean	Standar Deviasi	Maksimum		Minimum	
			Provinsi	Angka	Provinsi	Angka
X <sub>1</sub>	46,31	8,23	Bali	62,07	Kep. Riau	28,11
X <sub>2</sub>	1581930	2614940	Jawa Timur	9710661	Kep. Riau	324,01
X <sub>3</sub>	81945212	45629352	Riau	235350879	Papua	17412788
X <sub>4</sub>	59,35	10,25	Papua	84,43	Kep. Riau	33,67

Berdasarkan Tabel 3 dapat diketahui data menunjukkan bahwa pada variabel X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>, dan X<sub>4</sub> memiliki standar deviasi yang cukup besar, yang mencerminkan adanya variasi data yang signifikan di antara provinsi-provinsi. Dilihat dari nilai maksimum setiap variabel, diketahui bahwa tidak didominasi oleh Provinsi tertentu, sedangkan pada nilai minimum setiap variabel didominasi oleh Provinsi Kepulauan Riau. Hal ini dikarenakan kurang memadainya kondisi tanah yang cenderung berbukit dan tidak subur untuk pertanian, serta sumber daya air yang terbatas sehingga membuat sulitnya sektor pertanian untuk menjadi sumber pencaharian utama di Provinsi Kepulauan Riau.

##### Pengujian Asumsi Normal Multivariat

Uji distribusi normal multivariat dilakukan untuk mengetahui apakah suatu data multivariat berdistribusi normal multivariat atau tidak. Hasil pengujian asumsi distribusi normal multivariat pada data faktor-faktor sektor pertanian di Indonesia tahun 2023 adalah sebagai berikut.

Hipotesis:

$H_0$  : Data Faktor-Faktor Sektor Pertanian Di Indonesia Tahun 2023 memenuhi distribusi normal secara multivariat.

$H_1$  : Data Faktor-Faktor Sektor Pertanian Di Indonesia Tahun 2023 tidak memenuhi distribusi normal secara multivariat.

Ditetapkan taraf signifikan  $\alpha$  sebesar 0,05, sehingga didapatkan daerah penolakan tolak  $H_0$  jika nilai  $T$ -proporsi berada diluar  $45\% \leq T\text{-Proporsi} \leq 55\%$ .

Statistik uji:

Tabel 4. Uji Normal Multivariat

T-proporsi
0,529 atau 53%

Pada Tabel 4 di atas menunjukkan bahwa hasil uji distribusi normal multivariat memiliki T-proporsi sebesar 53% yang berada di dalam rentang 45% hingga 55%. Sehingga diputuskan gagal tolak  $H_0$  yang artinya data faktor-faktor sektor pertanian di Indonesia tahun 2023 telah berdistribusi normal multivariat.

### **Pengujian Asumsi Independensi**

Pengujian asumsi independensi bertujuan untuk mengetahui apakah data sama atau tidak sama dengan matriks identitas. Apabila matriks korelasi data tidak sama dengan matriks identitas, artinya asumsi dependensi terpenuhi. Pengujian *Bartlett* pada data faktor-faktor sektor pertanian di Indonesia tahun 2023 adalah sebagai berikut.

Hipotesis:

$H_0$  : Matriks korelasi identik dengan matriks identitas atau faktor-faktor sektor pertanian di Indonesia tahun 2023 independen.

$H_0$  : Matriks korelasi tidak identik dengan matriks identitas atau faktor-faktor sektor pertanian di Indonesia tahun 2023 dependen.

Ditetapkan taraf signifikan  $\alpha$  sebesar 0,05 maka didapatkan daerah penolakan yaitu tolak  $H_0$  ditolak jika nilai  $\chi^2_{hitung} > \chi^2_{(0.05;33)}$  atau  $P\text{-Value} < \alpha$ .

Statistik uji:

Tabel 5. Statistik Uji Dependensi

$\chi^2_{hitung}$	$\chi^2_{(0.05;33)}$	P-Value
25,258	47,339	0,000

Pada Tabel 4 menunjukkan bahwa hasil uji *Bartlett* didapatkan nilai  $\chi^2_{hitung}$  sebesar 25,258, di mana nilai tersebut lebih kecil dari  $\chi^2_{(0,05;33)}$  sebesar 47,339. Diperkuat dengan nilai *P-Value* sebesar 0,000 yang lebih kecil dari  $\alpha$  sebesar 0,05. Maka diputuskan tolak  $H_0$  yang artinya matriks korelasi tidak identik dengan matriks identitas atau faktor-faktor sektor pertanian di Indonesia tahun 2023 dependen.

### Pemilihan Kluster Optimum

Kluster optimum dipilih berdasarkan kriteria nilai *Pseudo-F* terbesar, selanjutnya metode terbaik dipilih berdasarkan nilai *Icdrate* yang paling kecil. Perbandingan nilai *Pseudo-F* dan *Icdrate* dari analisis kluster pada faktor-faktor sektor pertanian Indonesia tahun 2023 disajikan pada tabel berikut.

Tabel 6. Perbandingan *Pseudo-F* dan *Icdrate*

Metode	Jumlah Kluster	<i>Pseudo-F</i>	<i>Icdrate</i>
<i>Single Linkage</i>	2	0,494	0,952
	3	0,629	0,941
	4	0,710	0,934
	5	22,791	0,305
	2	13,002	0,435
<i>K-Means</i>	3	12,016	0,454
	4	14,011	0,416
	5	17,281	0,366

Tabel 6 menunjukkan nilai-nilai *Pseudo-F* dan *Icdrate* dari analisis kluster menggunakan dua metode berbeda dengan jumlah kluster sebanyak 2, 3, 4, dan 5 kluster. Berdasarkan nilai *Pseudo-F* dapat diketahui bahwa hasil pengelompokkan baik menggunakan *Single Linkage* maupun *K-Means* optimum pada jumlah kluster sebanyak 5 kluster. Hal ini dikarenakan kluster sebanyak 5 memiliki nilai *Pseudo-F* paling besar dibandingkan jumlah kluster lain dengan nilai berturut-turut yakni 22,791 dan 17,281. Selanjutnya, dari kedua metode dibandingkan nilai *Icdrate* yang diketahui bahwa metode *Single Linkage* memiliki *Icdrate* terkecil yakni 0,305. Dengan demikian, dapat disimpulkan bahwa metode terbaik dalam pengelompokkan faktor-faktor sektor pertanian Indonesia tahun 2023 yakni menggunakan metode *Single Linkage* dengan 5 kluster.

### Analisis Kluster Hierarki

Kluster hierarki digunakan untuk mengelompokkan objek yang memiliki kesamaan para karakteristik yang paling dekat, dengan menghitung nilai jarak *euclidean* antar kluster. Pada bab ini diuraikan hasil analisis kluster hierarki berupa hasil pengelompokkan, dendrogram, serta visualiasi dan karakteristik data dari hasil pengelompokkan faktor-faktor yang mempengaruhi sektor pertanian Indonesia tahun 2023.

## Pengelompokkan Menggunakan Metode Hirarki

Pengelompokkan provinsi di Indonesia berdasarkan faktor-faktor yang mempengaruhi sektor pertanian menggunakan analisis kluster hierarki terpilih dengan metode *Single Linkage* yang hasilnya disajikan pada tabel berikut.

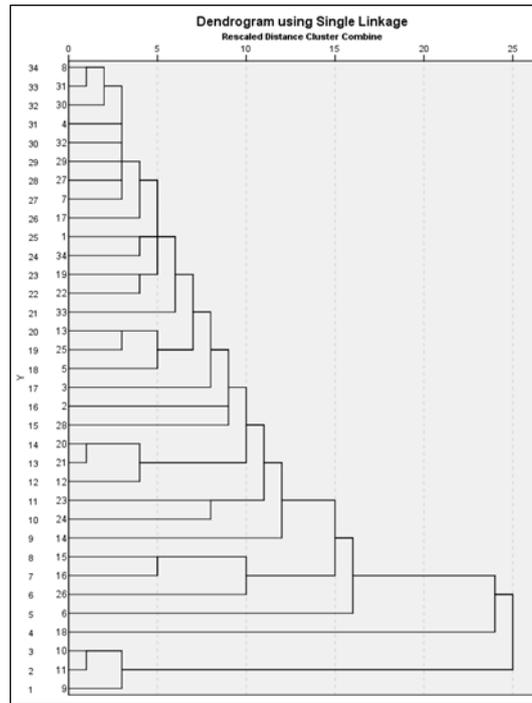
Tabel 7. Hasil Kluster Hierarki

Kluster	Jumlah	Provinsi
1	26	Aceh, Sumatera Barat, Sumatera Utara, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Sulawesi Barat, Gorontalo, Maluku, Maluku Utara, Papua, Papua Barat
2	1	DKI Jakarta,
3	3	Jawa Barat, Jawa Tengah, Jawa Timur
4	3	Riau, Kalimantan Timur, Kalimantan Selatan
5	1	Kapulauan Riau

Tabel di atas menunjukkan bahwa pengelompokkan provinsi di Indonesia berdasarkan faktor-faktor sektor pertanian tahun 2023 menggunakan metode *Single Linkage* optimum dengan jumlah kluster sebanyak 5. Berdasarkan tabel di atas dapat diketahui bahwa kluster yang memiliki jumlah anggota terbanyak yaitu pada kluster 1 dengan anggota sebanyak 26 provinsi. Selanjutnya, kluster 3 dan 4 memiliki jumlah anggota yang sama yakni sebanyak 3 provinsi. Sementara itu, kluster dengan anggota paling sedikit yakni kluster 1 dan 5 yang terdiri hanya 1 provinsi.

### Dendrogram

Metode *Single Linkage* mengelompokkan objek berdasarkan kedekatan jarak *euclidean* yang kemudian menghasilkan hierarki dan dapat divisualisasikan menggunakan dendrogram. Hierarki yang dihasilkan dari pengelompokkan provinsi di Indonesia berdasarkan faktor-faktor sektor pertanian menggunakan metode *Single Linkage* divisualisasikan dengan dendrogram pada gambar berikut.

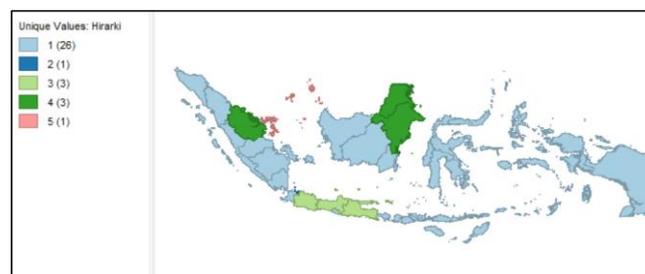


**Gambar 1.** Dendrogram *Single Linkage*

Dendrogram pada gambar di atas menunjukkan apabila ingin membentuk klaster sebanyak 5 klaster, maka klaster 1 akan memiliki anggota sebanyak 26 provinsi, klaster 2 sebanyak 3 provinsi, klaster 3 dan 4 memiliki anggota sebanyak 1 provinsi, dan klaster 5 memiliki anggota sebanyak 3 provinsi.

### Visualisasi Klaster Hierarki

Pengelompokkan provinsi di Indonesia berdasarkan faktor-faktor sektor pertanian menggunakan metode *Single Linkage* dapat divisualisasikan dengan peta tematik pada gambar berikut.



**Gambar 2.** Pemetaan Klaster Hierarki

Gambar 2 menunjukkan pemetaan provinsi Indonesia secara hierarki berdasarkan faktor-faktor sektor pertanian tahun 2023 yang terbagi menjadi 5 klaster. Berdasarkan gambar dapat diketahui bahwa klaster 1 ditunjukkan oleh warna biru muda yang mendominasi hampir seluruh pulau. Sementara itu, klaster 2 ditunjukkan dengan warna biru tua yang hanya berada di ujung barat Pulau Jawa. Klaster 3 ditunjukkan dengan warna hijau muda yang mendomuniasi Pulau Jawa. Klaster 4 ditunjukkan dengan warna hijau tua yang berada di Pulau Sumatera dan

Kalimantan. Terakhir, kluster 5 ditunjukkan dengan warna merah muda yang hanya ada pada Kepulauan dekat Sumatera.

### Karakteristik Data Berdasarkan Kluster Hirarki

Karakteristik data pada setiap kluster yang terbentuk menggunakan *Single Linkage* disajikan pada tabel berikut.

Tabel 8. Karakteristik Data Kluster Hierarki

Kluster	Variabel			
	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>
1	46.59	976587.72	69885599.72	62.36
2	49.26	2674.28	106087809.8	36.23
3	56.73	9311602.69	51022212.65	59.55
4	38.58	152182.21	198732547.6	49.39
5	28.11	324.01	113759531.7	33.67

Berdasarkan tabel di atas dapat diketahui bahwa karakteristik X<sub>1</sub> maupun X<sub>2</sub> yang memiliki rata-rata tertinggi adalah kluster 3. Nilai rata-rata kedua variabel yang tinggi menunjukkan bahwa kluster 3 memiliki tingkat produktivitas sektor pertanian yang tinggi, hal ini dikarenakan kluster 3 yang didominasi provinsi di Pulau Jawa memiliki kualitas lahan pertanian yang lebih baik. Sementara itu, karakteristik variabel X<sub>3</sub> yang memiliki rata-rata tertinggi adalah kluster 4, hal ini menunjukkan bahwa kluster 4 memiliki kondisi produktivitas pertanian cukup tinggi dengan jumlah tenaga kerja yang lebih sedikit, sehingga nilai tambah per tenaga kerjanya menjadi tinggi. Kemudian, karakteristik variabel X<sub>4</sub> yang memiliki rata-rata tertinggi adalah kluster 1, hal ini menunjukkan bahwa kluster 1 memiliki kondisi di mana lahan pertanian masih banyak yang belum dikelola sehingga lapangan pekerjaan di sektor pertanian menjadi tinggi.

### Analisis Kluster Non-Hierarki

Kluster non-hierarki algoritma *K-Means* digunakan untuk mengelompokkan objek berdasarkan kedekatan jarak *euclidean* antar objek dengan *centroid*. Pada bab ini diuraikan hasil analisis kluster non-hierarki berupa initial cluster, iteration history, final cluster center, hasil pengelompokkan, visualiasi dan karakteristik data berdasarkan kluster, serta hasil uji signifikansi kluster dari hasil pengelompokkan faktor-faktor yang mempengaruhi sektor pertanian Indonesia tahun 2023.

#### Initial Cluster

*Initial cluster* merupakan nilai *centroid* awal sebelum melakukan iterasi. *Initial cluster* dari pengelompokkan provinsi di Indonesia berdasarkan faktor-faktor sektor pertanian tahun 2023 disajikan pada tabel berikut.

Tabel 9. *Initial Cluster*

Variabel	Klaster				
	1	2	3	4	5
Z(X <sub>1</sub> )	-2,178	-0,794	1,302	-0,318	0,862
Z(X <sub>2</sub> )	-0,596	-0,518	3,063	-0,506	-0,016
Z(X <sub>3</sub> )	0,687	3,312	-0,785	0,890	-0,946
Z(X <sub>4</sub> )	-2,467	-0,694	0,361	-0,116	1,244

Tabel di atas menunjukkan nilai *centroid* awal pada masing-masing klaster untuk setiap variabel yang sudah distandarisasi. Berdasarkan tabel di atas dapat diketahui bahwa klaster yang memiliki rata-rata tertinggi pada variabel X<sub>1</sub> adalah klaster 3 dengan *centroid* awal sebesar 1,302. Klaster yang memiliki rata-rata tertinggi pada variabel X<sub>2</sub> adalah klaster 3 dengan *centroid* awal sebesar 3,068. Klaster yang memiliki rata-rata tertinggi pada variabel X<sub>3</sub> adalah klaster 2 dengan *centroid* awal sebesar 3,312. Klaster yang memiliki rata-rata tertinggi pada variabel X<sub>4</sub> adalah klaster 5 dengan *centroid* awal sebesar 1,244.

### Iteration History

Pada analisis klaster non-hierarki dengan algoritma *K-Means* perlu dilakukan iterasi untuk mendapatkan hasil pengelompokan yang tepat. *Iteration history* dari pengelompokan provinsi di Indonesia berdasarkan faktor-faktor sektor pertanian tahun 2023 disajikan pada tabel berikut.

Tabel 10. *Iteration History*

Iterasi	Klaster				
	1	2	3	4	5
1	0	0,844	0,395	0,862	0,719
2	0	0	0	0,107	0,133
3	0	0	0	0,151	0,127
4	0	0	0	0	0

Berdasarkan tabel di atas, dapat diketahui bahwa dalam pengelompokan provinsi di Indonesia berdasarkan faktor-faktor sektor pertanian tahun 2023 menggunakan *K-Means* dibutuhkan 4 kali proses iterasi sehingga didapatkan 5 klaster yang tepat. Adapun jarak minimum antar pusat klaster yang didapatkan setelah melakukan iterasi adalah 2,535.

### Final Cluster Center

Setelah melakukan iterasi, maka akan didapatkan nilai *centroid* akhir dari proses pengelompokan provinsi di Indonesia berdasarkan faktor-faktor sektor pertanian tahun 2023 menggunakan *K-Means* yang hasilnya disajikan pada tabel berikut.

Tabel 11. *Final Cluster Center*

Variabel	Klaster				
	1	2	3	4	5
Z(X <sub>1</sub> )	-2,178	-0,925	1,248	-0,498	0,549
Z(X <sub>2</sub> )	-0,596	-0,539	2,912	-0,442	-0,056
Z(X <sub>3</sub> )	0,687	2,522	-0,668	-0,083	-0,369
Z(X <sub>4</sub> )	-2,467	-0,957	0,019	-0,352	0,704

Tabel di atas menunjukkan nilai *centroid* akhir pada masing-masing klaster untuk setiap variabel yang didapatkan setelah melakukan iterasi. Berdasarkan tabel di atas dapat diketahui bahwa klaster yang memiliki rata-rata tertinggi pada variabel  $X_1$  adalah klaster 3 dengan *centroid* akhir sebesar 1,248. Klaster yang memiliki rata-rata tertinggi pada variabel  $X_2$  adalah klaster 3 dengan *centroid* akhir sebesar 2,912. Klaster yang memiliki rata-rata tertinggi pada variabel  $X_3$  adalah klaster 2 dengan *centroid* akhir sebesar 2,522. Klaster yang memiliki rata-rata tertinggi pada variabel  $X_4$  adalah klaster 5 dengan *centroid* akhir sebesar 0,704.

### **Pengelompokkan Klaster Non-Hirarki**

Pengelompokkan provinsi di Indonesia berdasarkan faktor-faktor yang mempengaruhi sektor pertanian menggunakan analisis klaster non-hierarki dengan algoritma *K-Means* yang hasilnya disajikan pada tabel berikut.

Tabel 12. Hasil Klaster Non-Hierarki

Klaster	Jumlah	Provinsi
1	1	Kepulauan Riau
2	3	Riau, Kalimantan Timur, Kalimantan Utara
3	3	Jawa Barat, Jawa Tengah, Jawa Timur
4	13	Jambi, Kepulauan Bangka Belitung, DKI Jakarta, DI Yogyakarta, Banten, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tenggara, Maluku, Maluku Utara, Papua Barat
5	14	Aceh, Sumatera Utara, Sumatera Barat, Sumatera Selatan, Bengkulu, Lampung, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Tengah, Sulawesi Selatan, Gorontalo, Sulawesi Barat, Papua

Tabel di atas menunjukkan bahwa pengelompokkan provinsi di Indonesia berdasarkan faktor-faktor sektor pertanian tahun 2023 menggunakan metode *K-Means* optimum dengan jumlah klaster sebanyak 5. Berdasarkan tabel di atas dapat diketahui bahwa klaster yang memiliki jumlah anggota terbanyak yaitu pada klaster 5 dengan anggota sebanyak 14 provinsi. Klaster dengan anggota terbanyak kedua adalah klaster 4 dengan anggota sebanyak 13 provinsi. Selanjutnya, klaster 2 dan 3 memiliki jumlah anggota yang sama yakni sebanyak 3 provinsi. Sementara itu, klaster dengan anggota paling sedikit yakni klaster 1 yang terdiri hanya 1 provinsi.

### **Visualisasi Klaster Non-Hierarki**

Pengelompokkan provinsi di Indonesia berdasarkan faktor-faktor sektor pertanian menggunakan metode *K-Means* dapat divisualisasikan dengan peta tematik pada gambar berikut.



**Gambar 3.** Pemetaan Klaster Non-Hierarki

Gambar 2 menunjukkan pemetaan provinsi Indonesia secara non-hierarki berdasarkan faktor-faktor sektor pertanian tahun 2023 yang terbagi menjadi 5 klaster. Berdasarkan gambar dapat diketahui bahwa klaster 1 ditunjukkan oleh warna biru muda yang hanya ada di kepulauan dekat Sumatera. Sementara itu, klaster 2 ditunjukkan dengan warna biru tua yang berada di Pulau Sumatera dan Kalimantan. Klaster 3 ditunjukkan dengan warna hijau tua yang mendominasi di Pulau Jawa. Klaster 4 ditunjukkan dengan warna hijau tua yang tersebar pada kelima pulau besar di Indonesia. Terakhir, klaster 5 ditunjukkan dengan warna merah muda yang hampir tersebar diseluruh wilayah kecuali di Pulau Kalimantan.

*Karakteristik Data Berdasarkan Klaster Non-Hirarki*

Karakteristik data pada setiap klaster yang terbentuk menggunakan *K-Means* disajikan pada tabel berikut.

Tabel 13. Karakteristik Data Klaster Non-Hierarki

Klaster	Variabel			
	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>
1	28.11	324.01	113759531.67	33.67
2	38.58	152182.21	198732547.60	49.39
3	56.73	9311602.69	51022212.65	59.55
4	42.15	409160.63	78087057.38	55.69
5	50.90	1433919.05	64855832.61	66.68

Berdasarkan tabel di atas dapat diketahui bahwa karakteristik X<sub>1</sub> maupun X<sub>2</sub> yang memiliki rata-rata tertinggi adalah klaster 3. Nilai rata-rata kedua variabel yang tinggi menunjukkan bahwa klaster 3 memiliki tingkat produktivitas sektor pertanian yang tinggi, hal ini dikarenakan klaster 3 yang didominasi provinsi di Pulau Jawa memiliki kualitas lahan pertanian yang lebih baik. Sementara itu, karakteristik variabel X<sub>3</sub> yang memiliki rata-rata tertinggi adalah klaster 2, hal ini menunjukkan bahwa klaster 2 memiliki kondisi produktivitas pertanian cukup tinggi dengan jumlah tenaga kerja yang lebih sedikit, sehingga nilai tambah per tenaga kerjanya menjadi tinggi. Kemudian, karakteristik variabel X<sub>4</sub> yang memiliki rata-rata tertinggi adalah klaster 5, hal ini menunjukkan bahwa klaster 5 memiliki kondisi di mana lahan pertanian masih banyak yang belum dikelola sehingga lapangan pekerjaan di sektor pertanian menjadi tinggi.

## Uji Signifikansi Klaster

Analisis ANOVA digunakan untuk melihat adanya perbedaan variabel pada klaster yang terbentuk. Hasil ANOVA dari pengelompokan provinsi di Indonesia berdasarkan faktor-faktor sektor pertanian tahun 2023 diuraikan sebagai berikut.

Hipotesis:

$H_0: \beta_j = 0; j = 1, 2, 3, 4, 5$  (Variabel prediktor tidak berpengaruh signifikan terhadap pengelompokan provinsi di Indonesia tahun 2023 berdasarkan faktor-faktor sektor pertanian)

$H_1$ : Minimal ada 1  $\beta_j \neq 0$  (Minimal terdapat satu variabel prediktor berpengaruh signifikan terhadap pengelompokan provinsi di Indonesia tahun 2023 berdasarkan faktor-faktor sektor pertanian)

Ditetapkan taraf signifikan ( $\alpha$ ) sebesar 0,05 sehingga didapatkan daerah penolakan yaitu tolak  $H_0$  jika  $F_{hitung} > F_{(0,05;4;29)}$  atau  $p\text{-value} < 0,05$ .

Statistik uji:

Tabel 14. ANOVA

Variabel	$F_{hitung}$	$F_{(0,05;4;29)}$	p-value	Keputusan
$\beta_1$	10,378	2,701	0,000	Tolak $H_0$
$\beta_2$	56,537		0,000	Tolak $H_0$
$\beta_3$	16,392		0,000	Tolak $H_0$
$\beta_4$	8,069		0,000	Tolak $H_0$

Tabel di atas menunjukkan bahwa semua variabel prediktor memiliki nilai  $F_{hitung}$  yang lebih besar dari  $F_{(0,05;1;29)}$  senilai 2,701 serta di dukung dengan nilai  $p\text{-value}$  yang lebih kecil dari  $\alpha$  sebesar 0,05 sehingga diputuskan tolak  $H_0$  pada semua variabel. Dengan demikian, semua variabel prediktor yang di analisis berpengaruh signifikan terhadap pengelompokan provinsi di Indonesia berdasarkan faktor-faktor sektor pertanian tahun 2023.

## 5. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis klaster yang telah dilakukan, dapat ditarik kesimpulan bahwa pengelompokan provinsi di Indonesia berdasarkan faktor-faktor yang berpengaruh pada sektor pertanian tahun 2023, baik secara hierarki dengan *Single Linkage* maupun secara non-hierarki dengan *K-Means* didapatkan klaster optimum yang terbentuk sebanyak 5 klaster. Hasil uji signifikansi klaster menunjukkan bahwa seluruh variabel yang digunakan dalam analisis berpengaruh signifikan dalam pengelompokan provinsi di Indonesia berdasarkan faktor-faktor yang berpengaruh pada sektor pertanian tahun 2023. Berdasarkan karakteristik variabel menunjukkan bahwa klaster yang anggotanya didominasi provinsi di Pulau Jawa memiliki produktivitas sektor pertanian yang tinggi. Kemudian, klaster yang beranggotakan provinsi Riau, Kalimantan Timur, dan Kalimantan Utara memiliki karakteristik produktivitas pertanian

cukup tinggi dengan jumlah tenaga kerja yang sedikit. Sementara itu, kluster dengan anggota yang paling banyak memiliki karakteristik banyak lahan pertanian yang belum dikelola sehingga lapangan pekerjaan di sektor pertanian menjadi tinggi.

## DAFTAR PUSTAKA

- Amah, N., Wahyuningsih, S., & Amijaya, F. D. T. (2017). Analisis cluster non-hierarki dengan menggunakan metode K-modes pada mahasiswa program studi statistika angkatan 2015 FMIPA Universitas Mulawarman. *Jurnal Eksponensial*, 8(1), 9-16.
- Baroroh, A. (2012). *Analisis multivariat dan time series dengan SPSS 21*. Jakarta: PT. Elex Media Komputindo.
- Fitriana, I. N. L. (2021). Pengelompokan provinsi di Indonesia berdasarkan indikator keluarga sehat menggunakan metode kluster hierarki dan non-hierarki. *Jurnal Paradigma: Jurnal Multidisipliner Mahasiswa Pascasarjana Indonesia*, 2(1), 27-36.
- Ghozali, I. (2018). *Aplikasi analisis multivariate dengan program IBM SPSS 25*. Semarang: Badan Penerbit Universitas Diponegoro.
- Hendrawaty, E. (2006). Pengembangan konsep manajemen mutu terpadu bagi badan usaha milik negara (BUMN) jasa keuangan cabang Bandarlampung. *Jurnal Bisnis dan Manajemen*, 2(3), 159-256.
- Irwansyah, E., & Faisal, M. (2015). *Advanced Clustering: Teori dan aplikasi*. Yogyakarta: DEEPUBLISH.
- Johnson, R. A., & Wichern, D. W. (2014). *Applied multivariate statistical analysis* (6th ed.). New York: Pearson Prentice Hall.
- Kusumaningrum, S. I. (2019). Pemanfaatan sektor pertanian sebagai penunjang pertumbuhan perekonomian Indonesia. *Jurnal Transaksi*, 11(1), 80-89.
- Ningrat, D. R., Maruddani, D. A. I., & Wuryandari, T. (2016). Analisis cluster dengan algoritma K-means dan fuzzy C-means clustering untuk pengelompokan data obligasi korporasi. *Jurnal Gaussian*, 5(4).
- Rizki. (2020). *Pertanian dan ketenagakerjaan: Analisis lapangan kerja informal di Indonesia*. Yogyakarta: Andi.
- Robert, K. (2019). *Cara mudah belajar statistik analisis data dan eksplorasi*. Jakarta: Penerbit Kencana.
- Simamora, B. (2005). *Analisis multivariat pemasaran*. Jakarta: PT. Gramedia Pustaka Utama.
- Siregar, S. (2017). *Statistika terapan untuk perguruan tinggi*. Jakarta: Kencana.
- Snapp, S. (2017). *Agricultural systems: Agroecology and rural innovation for development* (2nd ed.). ResearchGate.
- Soetomo. (2019). *Dampak tenaga kerja informal terhadap pertanian*. Yogyakarta: UGM.

Sudaryono. (2021). *Statistik II: Statistik inferensial untuk penelitian*.

Surya. (2018). *Ekonomi pertanian: Teori dan aplikasi*. Jakarta: Rajawali Pers.

Wahyudi, K. D. (2012). Kebijakan strategis usaha pertanian dalam rangka peningkatan produksi dan pengentasan kemiskinan. *Majalah Ilmiah "DIAN ILMU"*, 11(2), 78-91.

Windasari, R. (2020). *Analisis cluster hierarki metode average linkage berdasarkan jumlah kriminalitas di Indonesia tahun 2019*. Malang: Repository UIN Maulana Malik Ibrahim Malang.