

**PENGELOMPOKAN PROVINSI-PROVINSI DI INDONESIA  
MENGUNAKAN METODE WARD  
(Studi Kasus: Produksi Tanaman Pangan di Indonesia Tahun 2018)**

**Besya Salsabilla Azani Arif<sup>1</sup>, Agus Rusgiyono<sup>2</sup>, Abdul Hoyyi<sup>3</sup>**

<sup>1,2,3</sup>Departemen Statistika FSM Universitas Diponegoro

Email: [besyaoca2425@gmail.com](mailto:besyaoca2425@gmail.com)

**ABSTRACT**

Cluster analysis is a technique for grouping objects or observations into homogeneous groups. Cluster analysis is divided into two methods, namely hierarchy and non-hierarchy. The hierarchy method generally involves a series of  $n-1$  decisions ( $n$  is the number of observations) that combine observations into a tree-like structure or dendrogram. Hierarchy is divided into two methods, namely agglomerative (concentration) and splitting (distribution). For non-hierarchical methods, the number of clusters can be determined by the researcher. Ward method is a hierarchical cluster analysis method that can maximize homogeneity in the cluster. The *Sum-of-Square* (SSE) formula is used in this method to minimize variations in the clusters that are formed. In this research, squared euclid distance is used to measure the similarity between object pairs. The data used in this study are secondary data on food crop production, namely rice, corn, soybeans, peanuts, green beans, sweet potatoes, and cassava in Indonesia 2018. To determine the cluster, the elbow method is used to form optimal clusters using WSS formula. Based on the analysis results, it was found that the optimal cluster is four clusters. The first cluster consists of 9 Province, the second cluster consists of 20 Province, the third cluster consists of 1 Province, the fourth cluster consists of 2 Province, and the fifth cluster consists of 2 Province.

**Keywords:** Food Crop, Cluster Analysis, Ward Method, Squared Euclid, Elbow Method

**1. PENDAHULUAN**

Indonesia termasuk negara agraris yaitu profesi sebagian besar penduduk Indonesia adalah petani. Sehingga sebagian besar perekonomian di Indonesia ditopang oleh petani. Menurut data Badan Pusat Statistik (BPS) tahun 2014, sebagian luas lahan yang ada di Indonesia yaitu 71,33 persen digunakan untuk usaha pertanian. Pada tahun 2014 luas lahan yang digunakan untuk usaha pertanian mencapai 36,83 juta hektar<sup>[10]</sup>

Manusia dengan segala keterbatasannya selalu berusaha dan ingin memenuhi kebutuhannya yang tentunya berkualitas dan terkecukupi. Karena keinginan itu, suatu hal utama yang harus dicukupi adalah kebutuhan pangan. Undang Undang No 7 Tahun 1996 tentang pangan menyatakan bahwa tujuan pangan yaitu mencapai kecukupan pangan. Untuk mencapai tujuan tersebut, kebijakan pangan dilakukan

guna menjamin ketersediaan pangan setiap saat dalam jumlah yang cukup, merata, aman, bermutu, bergizi, beragam, dan dapat dijangkau oleh daya beli masyarakat. Analisis kluster adalah suatu kelas teknik, dipergunakan untuk mengklasifikasikan objek atau kasus (responden) ke dalam kelompok yang relatif homogen, yang disebut klaster<sup>[12]</sup>. Teknik kluster merupakan bagian dari multivariat. Konsep variat yang digunakan berbeda dari konsep variat pada teknik-teknik lainnya. Pada teknik kluster variat dianggap sebagai sejumlah variabel (dianggap sebagai karakteristik) yang dipakai untuk membandingkan sebuah obyek dengan obyek lainnya. Salah satu metode varians yang umum adalah metode *ward*. Dimana karakteristik dari Metode Ward yaitu mencari rata-rata di masing masing klaster<sup>[4]</sup>.Maka dari itu, penulis memutuskan untuk melakukan penelitian menggunakan pengelompokan dengan metode ward dengan variabel jenis-jenis tanaman pangan yang terdiri dari padi, jagung, kedelai, kacang hijau, kacang tanah, ubi kayu, dan ubi jalar di Provinsi Indonesia Tahun 2018.

## 2. TINJAUAN PUSTAKA

### 2.1. Tanaman Pangan

Tanaman Pangan yaitu semua model tanaman yang di dalamnya ada karbohidrat serta protein sebagai sumber energi manusia. Tanaman pangan adalah tanaman paling utama yang dikonsumsi manusia untuk badan. Umumnya tanaman pangan yaitu tanaman yang tumbuh jangka waktu semusim. Di Indonesia, ada 7 jenis tanaman pangan yang dibutuhkan sehari-hari bahkan menjadi bahan baku utama, yaitu<sup>[6]</sup>:

1. Padi.
2. Jagung.
3. Kedelai.
4. Kacang Tanah.
5. Kacang Hijau.
6. Ubi Jalar
7. Ubi Kayu.

### 2.2. Analisis Kluster

Analisis kluster merupakan suatu teknik, yang dipergunakan untuk mengklasifikasi atau mengelompokkan objek atau kasus (responden) ke dalam kelompok yang relatif homogen. Di dalam analisis kluster, hubungan interdependensi antara seluruh set variabel dikaji. Metode ini tidak ada pembeda antara variabel bebas dan variabel tidak bebas. Objek atau kasus dalam tiap kelompok cenderung mirip satu sama lain dan berbeda jauh dengan objek dari kluster lainnya<sup>[12]</sup>.

### 2.3. Metode Ward

Metode Ward adalah metode clustering hirarki yang bersifat agglomerative untuk memperoleh cluster yang memiliki varian internal sekecil mungkin. Agglomerative merupakan prosedur pengelompokan hirarki dimana setiap objek berawal dari cluster yang terpisah. Cluster-cluster dibentuk dengan mengelompokkan objek ke dalam cluster yang semakin banyak objek yang menjadi anggotanya. Proses ini dilanjutkan sampai semua objek menjadi anggota dari cluster tunggal<sup>[12]</sup>. Menurut Rencher Jika AB adalah kluster yang diperoleh dari menggabungkan kluster A dan B, maka jumlah perbedaan *witihin-cluster* (vektor dari rata-rata kluster) adalah<sup>[11]</sup>:

$$SSE_A = \sum_{i=1}^{n(A)} (\mathbf{y}_i - \bar{\mathbf{y}}_A)' (\mathbf{y}_i - \bar{\mathbf{y}}_A) \quad (1)$$

$$SSE_B = \sum_{i=1}^{n(B)} (\mathbf{y}_i - \bar{\mathbf{y}}_B)' (\mathbf{y}_i - \bar{\mathbf{y}}_B) \quad (2)$$

$$SSE_{AB} = \sum_{i=1}^{n(AB)} (\mathbf{y}_i - \bar{\mathbf{y}}_{AB})' (\mathbf{y}_i - \bar{\mathbf{y}}_{AB}) \quad (3)$$

Dengan  $n_A$ ,  $n_B$ , dan  $n_{AB} = n_A + n_B$  adalah jumlah titik dalam A, B, dan AB. Karena jumlah ini, perbedaan setara dengan jumlah *within-cluster* dalam kuadrat, dilambangkan dengan  $SSE_A$ ,  $SSE_B$ , dan  $SSE_{AB}$ .

Metode ward menggabungkan kluster A dan B yang dapat meminimalkan peningkatan SSE yang didefinisikan sebagai berikut:

$$I_{AB} = SSE_{AB} - (SSE_A + SSE_B) \quad (4)$$

Dapat ditunjukkan bahwa peningkatan  $I_{AB}$  bisa diuraikan sebagai berikut:

$$\begin{aligned} I_{AB} &= n_A (\bar{\mathbf{y}}_A - \bar{\mathbf{y}}_{AB})' (\bar{\mathbf{y}}_A - \bar{\mathbf{y}}_{AB}) + n_B (\bar{\mathbf{y}}_B - \bar{\mathbf{y}}_{AB})' (\bar{\mathbf{y}}_B - \bar{\mathbf{y}}_{AB}) \\ &= \frac{n_A n_B}{n_A + n_B} (\bar{\mathbf{y}}_A - \bar{\mathbf{y}}_B)' (\bar{\mathbf{y}}_A - \bar{\mathbf{y}}_B) \end{aligned} \quad (5)$$

Jadi dengan persamaan diatas, meminimalkan peningkatan SSE sama dengan meminimalkan perbedaan *within-cluster*. Jika A hanya terdiri dari  $\mathbf{y}_i$  dan B hanya terdiri dari  $\mathbf{y}_j$  maka  $SSE_A$  dan  $SSE_B$  sama dengan nol. Untuk persamaan (4) dan (5) menjadi

$$\begin{aligned} I_{ij} = SSE_{ij} &= \frac{1}{2} (\mathbf{y}_i - \mathbf{y}_j)' (\mathbf{y}_i - \mathbf{y}_j) \\ SSE_{ij} &= \frac{1}{2} \mathbf{d}^2(\mathbf{y}_i, \mathbf{y}_j) \end{aligned} \quad (6)$$

## 2.4. Pemilihan Ukuran Jarak

Dalam mengelompokkan objek dalam analisis kluster diperlukan ukuran untuk mengukur kedekatan dari objek satu ke objek lain agar menjadi suatu kluster. Untuk mengukur kedekatan diperlukan ukuran perbedaan<sup>[5]</sup>. Salah satu ukuran jarak yang bisa digunakan adalah kuadrat euclid (*squared euclid*). *Kuadrat euclid* Jumlah kuadrat untuk melihat perbedaan nilai pada tiap variabel tanpa mengakarkan dan membuat penghitungan akan lebih cepat. Jarak ini direkomendasikan untuk metode Ward dan metode centroid<sup>[4]</sup>. Kuadrat euclid di definisikan sebagai berikut:

$$d_{ij}^2 = \sum_{k=1}^p (x_{ik} - x_{jk})^2$$

dengan :

$d_{ij}^2$  : Kuadrat euclid antara objek ke- $i$  dan objek ke- $j$

$p$  : Jumlah variabel kluster

$x_{ik}$  : Data objek ke- $i$  pada variabel ke- $k$

$x_{jk}$  : Data objek ke- $j$  pada variabel ke- $k$

## 2.5. Asumsi Analisis Kluster

Terdapat 2 asumsi analisis kluster yang harus terpenuhi, yaitu:

### a. Sampel Mewakili Populasi

Sampel yang mewakili populasi atau sampel representative adalah sampel yang dapat dikatakan mewakili populasi yang ada. Pengujian sampel yang mewakili (*sample representative*) dapat dilakukan dengan uji KMO (*Kaiser Mayer Olkin*). Uji KMO digunakan untuk melihat syarat kecukupan suatu sampel. Uji KMO ini mengukur kecukupan sampling secara menyeluruh dan mengukur kecukupan *sampling* untuk setiap variabel. Sampel dikatakan dapat mewakili populasi atau

sampel representatif apabila diperoleh nilai KMO berkisar antara 0,5 sampai dengan 1<sup>[13]</sup>

b. Deteksi Multikoliniertias

Dalam analisis klaster, permasalahan multikolinieritas adalah adanya hubungan linier pada variabel bebas. Multikolinieritas bertindak sebagai pembobotan yang tidak terlihat oleh peneliti tapi mempengaruhi prosesnya analisis. Peneliti disarankan untuk menguji variabel-variabel mana saja yang mempunyai multikolinieritas<sup>[4]</sup>. Untuk mengetahui ada tidaknya multikolinieritas dapat digunakan nilai *Variance Inflation Factor* (VIF) yaitu<sup>[3]</sup>:

$$VIF_k = \frac{1}{1 - R_k^2}$$

dengan:

$VIF_k$  : *Variance Inflation Factor* variabel  $k$ ,  $k = 1, 2, \dots, p$

$R_k^2$  : Koefisiendeterminasi yang diperoleh bil nilai variabel ke  $k$

Jika nilai  $VIF \leq 10$  maka bisa disimpulkan bahwa tingkat multikolinieritas tidak tinggi atau masih bisa ditolerir.

**2.6. Pemilihan Jumlah Klaster Optimal**

Dalam analisis klaster, peneliti perlu menentukan klaster. Peneliti mengharapkan mendapatkan klaster yang paling baik dan optimal. Untuk menentukan klaster yang paling optimal diperlukan suatu metode yaitu metode elbow. Metode elbow digunakan untuk menentukan jumlah klaster yang optimal dari data. Metode ini merupakan metode visual. Idennya adalah dimulai dengan  $k = 2$ , dan terus meningkat dalam setiap langkah dengan ditambah 1 pada nilai  $k$ . Contoh pada nilai  $k=3$ , apabila terjadi perubahan drastis yang berbanding terbalik dengan nilai sebelumnya, maka nilai sebelum terjadinya perubahan tersebut dianggap sebagai jumlah klaster yang paling tepat. Menurut Malik & Tuckfield (2019) untuk mengidentifikasi klaster dalam data, peneliti mencoba meminimalkan jarak antara titik di klaster, dan *Within-Sum-of-Squares* (WSS) bisa mengukur hal itu. WSS adalah nilai jumlah kuadrat dari jarak semua titik dalam sebuah klaster. Berikut merupakan rumus WSS<sup>[7]</sup>

$$WSS = \sum_{l=1}^k \sum_{i \in S_k} \sum_{j=1}^p (X_{lj} - \bar{X}_{kj})^2$$

dengan:

$S_k$  : Pengamatan di klaster  $k$

$\bar{X}_{kj}$  : Rata-rata objek  $j$  dari klaster pusat untuk klaster  $k$

$p$  : Jumlah total dimensi data

**3. METODOLOGI PENELITIAN**

Data yang digunakan dalam tugas akhir ini merupakan data sekunder yaitu data produksi tanaman pangan se-Provinsi Indonesia Tahun 2018 diambil dari Kementerian Pertanian Republik Indonesia. Variabel yang digunakan dalam penelitian ini yaitu:

**Tabel 1.** Jenis-Jenis Tanaman Pangan di Indonesia

Variabel	Keterangan
X <sub>1</sub>	Padi
X <sub>2</sub>	Jagung
X <sub>3</sub>	Kedelai
X <sub>4</sub>	Kacang Tanah
X <sub>5</sub>	KacangHijau
X <sub>6</sub>	UbiJalar
X <sub>7</sub>	UbiKayu

Untuk obyek yang digunakan dalam penelitian ini adalah seluruh Provinsi di Indonesia yaitu berjumlah 34 Provinsi

*Software* yang digunakan sebagai alat hitung adalah *Microsoft Excel 2013* dan *Rx64*

3.6.1. Adapun langkah-langkah dalam analisis data adalah sebagai berikut:

1. *Input* data.
2. Menguji asumsi pada analisis kluster yaitu sampel mewakili populasi dan deteksi multikolinieritas.
3. Menghitung menggunakan jarak kuadrat euclid.
4. Melakukan pengklasteran menggunakan metode ward.
5. Menentukan kluster optimal menggunakan metode elbow.
6. Interpretasi dan pembuatan profil kluster menggunakan centroid.

## 4. ANALISIS DAN PEMBAHASAN

### 4.1 Asumsi Analisis Kluster

Terdapat dua asumsi pada analisis kluster yaitu sampel mewakili populasi dan deteksi multikolinieritas.

#### 4.1.1 Sampel Mewakili Populasi

Dilakukan dengan menggunakan *Kaiser Mayer Olkin* (KMO). Jika nilai KMO berkisar antara 0,5 sampai dengan 1 maka sampel mewakili populasi atau sampel representatif

**Tabel 2.** Nilai KMO

VIF	Keterangan
0,7005829	$0,5 \leq KMO \leq 1$

Berdasarkan Tabel 2, diperoleh nilai KMO sebesar 0,7005829. nilai tersebut berkisar antara 0,5 sampai dengan 1. Jadi, dapat disimpulkan bahwa sampel mewakili populasi atau sampel representatif.

#### 4.2.1 Deteksi Multikolinieritas

Untuk mendeteksi adanya multikolinieritas perlu dilakukan uji multikolinieritas dengan *Varians Inflation Factor* (VIF) . Variabel tidak mempunyai multikolinieritas yang tinggi jika nilai  $VIF \leq 10^{[1]}$ .

**Tabel 3.** Nilai VIF pada masing-masing Variabel

Variabel	VIF	Keterangan
X <sub>1</sub>	4,819605	VIF ≤10
X <sub>2</sub>	2,872154	VIF ≤10
X <sub>3</sub>	7,009755	VIF ≤10
X <sub>4</sub>	2,691966	VIF ≤10
X <sub>5</sub>	4,922002	VIF ≤10
X <sub>6</sub>	3,116263	VIF ≤10
X <sub>7</sub>	2,154895	VIF ≤10

Berdasarkan Tabel 3, Seluruh variabel mendapat nilai VIF ≤10. Maka dapat disimpulkan bahwa seluruh variabel tidak mempunyai multikolinieritas yang tinggi, maka bisa dilakukan analisis selanjutnya.

#### 4.2 Menghitung Perbedaan Hasil Produksi Tanaman Pangan Menggunakan Kuadrat Euclid

Berikut adalah langkah perhitungan perbedaan hasil produksi tanaman pangan antara Kabupaten-Kabupaten yang terpilih menggunakan kuadrat euclid yaitu<sup>[5]</sup>

$$d_{ij}^2 = \sum_{k=1}^p (x_{ik} - x_{jk})^2$$

Berikut adalah contoh perhitungan perbedaan hasil produksi tanaman pangan antar provinsi yang terpilih menggunakan kuadrat euclid

1. Kuadrat perbedaan hasil produksi tanaman pangan dari Provinsi Aceh dengan Provinsi Sumatera Utara yaitu

$$d_{12} = (1690554 - 1888836)^2 + (347324 - 1782714)^2 + (12831,2 - 32641,2)^2 + (3499,4807 - 4271,2522)^2 + (481,0248 - 1775,4017)^2 + (481,0248 - 81496)^2 + (18997,3 - 771485)^2$$

$$= 5,151811 \times 10^{12}$$

Perbedaan hasil produksi tanaman pangan dari Provinsi Aceh dengan Provinsi Sumatera Utara menggunakan kuadrat euclid.

2. Kuadrat perbedaan hasil produksi tanaman pangan dari Provinsi Aceh dengan Provinsi Sumatera Barat yaitu  $5,151811 \times 10^{12}$

$$d_{12} = (1690554 - 1509337)^2 + (347324 - 976956)^2 + (12831,2 - 2225,69)^2 + (3499,4807 - 4813,9814)^2 + (481,0248 - 311,5252)^2 + (481,0248 - 137319)^2 + (18997,3 - 203218)^2$$

$$= 4,810004 \times 10^{11}$$

Perbedaan hasil produksi tanaman pangan dari Provinsi Aceh dengan Provinsi Sumatera Barat menggunakan kuadrat euclid yaitu  $4,810004 \times 10^{11}$

Perhitungan kuadrat euclid terus dilakukan seperti pada point 1 dan 2. Karena terdapat 34 objek maka matriks yang terbentuk adalah matriks  $34 \times 34$

### 4.3 Pengklasteran Menggunakan Metode Ward

- i. Untuk proses pengklasteran metode ward, terlebih dahulu menghitung *sum squared error* (SSE), dengan rumus<sup>[11]</sup>

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

Berikut adalah contoh perhitungan *Sum Square error* (SSE) antar Provinsi

1. Perhitungan SSE antar dua objek yaitu Provinsi Aceh dengan Provinsi Sumatera Utara

$$\begin{aligned} SSE_{12} &= \frac{1}{2} \left\{ (1690554 - 1888836)^2 + (347324 - 1782714)^2 + (12831,2 - 32641,5)^2 \right. \\ &\quad + (3499,4807 - 4271,2522)^2 + (481,0248 - 1775,4017)^2 \\ &\quad \left. + (4380,78 - 81496)^2 + (18997,3 - 771485)^2 \right\} \\ &= 2,575905 \times 10^{12} \end{aligned}$$

2. SSE antar dua objek yaitu Provinsi Aceh dengan Provinsi Sumatera Barat

$$\begin{aligned} SSE_{12} &= \frac{1}{2} \left\{ (1690554 - 1509337)^2 + (347324 - 976956)^2 + (12831,2 - 2225,69)^2 \right. \\ &\quad + (3499,4807 - 4813,9814)^2 + (481,0248 - 311,5252)^2 \\ &\quad \left. + (4380,78 - 137319)^2 + (18997,3 - 203218)^2 \right\} \\ &= 2,405002 \times 10^{11} \end{aligned}$$

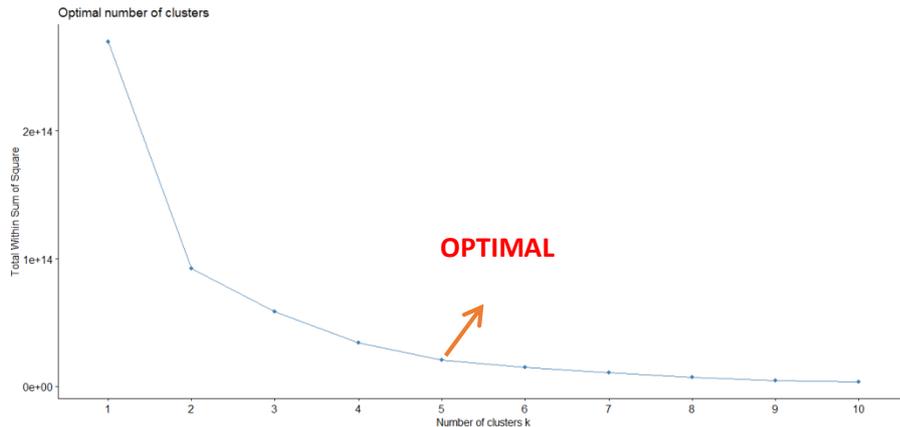
Dilakukan perhitungan seperti pada point 1 dan 2 sehingga membentuk matriks ukuran  $34 \times 34$ .

### 4.4 Pemilihan Jumlah Cluster Optimal

Untuk menentukan klaster yang optimal maka peneliti bisa menggunakan metode elbow. Dalam metode elbow, peneliti menentukan klaster dengan  $K=1$  sampai dengan  $K=10$ . Masing-masing klaster dihitung nilai *Within-Cluster Sum of Square* (WSS), total nilai WSS masing-masing  $K$  diperoleh sebagai berikut<sup>[9]</sup>

**Tabel 4.** Nilai WSS masing-masing  $K$

Klaster	WSS	Selisih
1	$2,696897 \times 10^{14}$	-
2	$9,215145 \times 10^{13}$	$1,77538 \times 10^{14}$
3	$5,882859 \times 10^{12}$	$3,33229 \times 10^{13}$
4	$3,442478 \times 10^{12}$	$2,44038 \times 10^{13}$
5	$2,076335 \times 10^{12}$	$1,336614 \times 10^{13}$
6	$1,478233 \times 10^{12}$	$5,98102 \times 10^{12}$
7	$1,084710 \times 10^{12}$	$3,93523 \times 10^{12}$
8	$7,331161 \times 10^{12}$	$3,51594 \times 10^{12}$
9	$4,491726 \times 10^{12}$	$2,83944 \times 10^{12}$



**Gambar 1.** Plot Metode Elbow

Dari Tabel 4, bahwa selisih nilai WSS paling sedikit dimulai pada  $K=5$ . Secara visual dari gambar 1, distorsi menurun secara lambat dari  $K=5$ . Maka dapat disimpulkan bahwa kluster optimal ada pada  $K=5$ .

#### 4.5 Interpretasi dan Pembuatan Profil Kluster

Untuk mengetahui karakteristik pada kluster maka diperlukan rata-rata tiap kluster yang terbentuk atau bisa disebut dengan centroid. Untuk hasil nilai centroid ada pada Tabel 5.

**Tabel 5.** Nilai Centroid masing-masing Kluster untuk Jumlah Kluster Lima

	Variabel						
	X1 Padi	X2 Jagung	X3 Kedelai	X4 Kacang Tanah	X5 Kacang Hijau	X6 Ubi Jalar	X7 Ubi Kayu
1	2.202.445,0	1.123.408,0	39.337,5	19.659,7	3.571,4	58.709,5	432.411,5
2	445.973,1	387.036,3	8.305,5	6.584,9	330,8	25.328,4	143.395,1
3	1.904.039,0	2.403.851,0	71.232,1	3.128,4	1.306,2	15.019,1	5.518.739,0
4	4.918.956,0	2.450.609,0	186.647,1	65.046,6	72.489,5	296.722,3	2.213.033,0
5	3.616.192,0	2.882.275,0	49.838,5	23.189,2	13.594,8	39.472,1	137.704,8

Keterangan:

- █ : Rata-rata Tertinggi
- █ : Rata-rata Terendah

Dari hasil Lima pengklasteran, dapat menunjukkan bahwa wilayah pada kluster empat yaitu Provinsi Jawa Barat dan Jawa Tengah merupakan Provinsi dengan hasil produksi padi, kedelai, kacang tanah, kacang hijau, dan ubi jalar tertinggi. Pada Variabel Produksi Ubi Kayu tertinggi ada pada kluster tiga yaitu

pada Provinsi Lampung. Pada kluster lima yaitu Provinsi NTB dan Sulawesi Selatan memiliki produksi jagung paling tinggi.

Pada kluster dua merupakan kluster yang memiliki banyak variabel dengan penghasil produksi tanaman pangan terendah yaitu produksi padi, jagung, kedelai, dan kacang hijau. Salah satu contoh wilayah dengan produksi terendah ada di Provinsi DKI Jakarta. Pada variabel produksi kacang tanah dan ubi jalar terendah berada di kluster tiga yaitu di Provinsi Lampung. Untuk Produksi Ubi Kayu terendah ada di Provinsi NTB dan Sulawesi Selatan.

Berikut adalah wilayah wilayah dari hasil pengklasteran lima kluster :

**Tabel 6.** Hasil Kluster Metode Ward untuk Jumlah Kluster Lima

<b>Kluster</b>	<b>Jumlah Anggota</b>	<b>Anggota Kluster</b>
1	9	Aceh, Sumatera Barat, Jawa Timur, Banten, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Tengah.
2	20	Sumatera Utara, Riau, Jambi, Sumatera Selatan, Bengkulu, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Yogyakarta, Bali, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua
3	1	Lampung
4	2	Jawa Barat, Jawa Tengah
5	2	NTB, Sulawesi Selatan

## 5. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah diuraikan pada bab sebelumnya, maka diperoleh kesimpulan sebagai berikut:

1. Berdasarkan hasil analisis menggunakan metode ward, diperoleh hasil pengklasteran optimal yaitu lima kluster.
2. Lima kluster yang terbentuk terdiri dari kluster pertama 9 Provinsi, kluster kedua 20 Provinsi, kluster tiga 1 Provinsi, kluster keempat 2 Provinsi, dan kluster kelima 2 Provinsi.
3. Kluster terbaik yaitu ada pada kluster 4 yaitu Jawa Tengah dan Jawa Barat memiliki produksi padi, kedelai, kacang tanah, kacang hijau, dan ubi jalar tertinggi. Untuk produksi Ubi Kayu tertinggi ada pada kluster 3 dan produksi Jagung tertinggi ada pada kluster 5.

Dari Point 3 bisa menjadi bahan evaluasi pemerintah terutama Dinas Pertanian untuk meningkatkan kualitas dan kuantitas tanaman pangan di Indonesia.

## 6. DAFTAR PUSTAKA

- [1] Ghozali, I. 2011. *Aplikasi Analisis Multivariate Dengan Program IBM SPSS 19*. Semarang: Badan Penerbitan Universitas Diponegoro
- [2] Gudono. 2014. *Analisis Data Multivariat*. Yogyakarta: BPF.
- [3] Gujarati, D. N. 1978. *Basic Econometrics*. New Jersey: The McGraw-Hill Companies
- [4] Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. 2010. *Multivariate Data Analysis*. New Jersey: Pearson Education, Inc.
- [5] Johnson, R. A., & Wichern, D. W. 2003. *Applied Multivariate Statistical Analysis*. New Jersey: Pearson Education, Inc.
- [6] [Kemenpan RI]. KementerianPertanianRepublik Indonesia.2019. <https://www.pertanian.go.id/home/?show=page&act=view&id=61>
- [7] Kodinariya, T. M., & Makwana, P. R. 2013. *Review on determining number of Cluster in K-Means Clustering*. International Journal of Advance Research in Computer Science and Management Studies
- [8] Kumari, V. V., Raju, B. R., & Naik, A. 2012. *Hybrid Clustering Algorithm based on Mahalanobis Distance and MST*. International Journal of Applied Information Systems (IJ AIS).
- [9] Malik, A., & Tuckfield, B. 2019. *Applied Unsupervised learning With R*. In Uncover hidden relationships and patterns with k-means clustering, hierarchical clustering, and PCA.
- [10] Noorjenah, Subagya, E. H., Iswadi, Amalia, R. R., Siagian, S. H., Poerwaningsih, R., Anggraeny, R. 2014. *Produksi Tanaman Pangan 2014*. Jakarta: Badan Pusat Statistik
- [11] Rencher, A. C. 2004. *Methods of Multivariate Analysis*. Canada: John Wiley & Sons
- [12] Supranto, J. 2018. *Analisis multivariat : Arti dan Interpretasi*. Jakarta: PT Rineka Cipta.
- [13] Widarjono, A. 2010. *Analisis Statistika Multivariat Terapan*. Yogyakarta: UPP STIM YKP