

# ANALISIS KLASTER DENGAN MENGGUNAKAN METODE *SINGLE LINKAGE* DAN METODE *K-MEANS*

(Studi Kasus: Data Produksi, Produktivitas, dan Luas Panen Tanaman Sayuran di  
Provinsi Bengkulu Tahun 2013)

Rezki Wahyuni<sup>1</sup>, Sigit Nugroho<sup>2</sup>, dan Pepi Novianti<sup>2</sup>

<sup>1</sup>Mahasiswa Jurusan Statistika, Prodi Matematika, FMIPA, UNIB, Bengkulu

<sup>2</sup>Dosen Pembimbing Jurusan Statistika, Prodi Matematika, FMIPA, UNIB, Bengkulu

[Rizqhywahyuni@gmail.com](mailto:Rizqhywahyuni@gmail.com)

## ABSTRAK

Analisis kluster merupakan teknik dalam analisis multivariat yang mempunyai tujuan untuk mengelompokkan obyek-obyek pengamatan menjadi beberapa kelompok berdasarkan karakteristik yang dimilikinya. Analisis kluster terdiri dari kluster hierarki dan kluster non hierarki. Salah satu metode pada kluster hierarki adalah kluster *Single Linkage*, dan kluster non hierarki adalah *K-Means*. Penelitian ini bertujuan untuk mengetahui hasil pengklasteran di Provinsi Bengkulu per Kabupaten, dan untuk mengetahui metode yang baik dari hasil pengklasteran metode *Single Linkage* dan metode *K-Means*, berdasarkan Data Produksi, Produktivitas, dan Luas Panen Tanaman Sayuran di Provinsi Bengkulu Tahun 2013. Penelitian ini menggunakan jarak *Euclid*. Berdasarkan hasil penelitian diperoleh pengelompokan data menjadi tiga anggota kluster yang terbentuk, dan memiliki hasil yang berbeda-beda dari setiap data dan metode yang digunakan. Dari hasil pengklasteran menyatakan bahwa metode *Single Linkage* merupakan metode yang baik dan Kabupaten Rejang Lebong memiliki pengaruh yang baik dengan variabel Cabe.

**Kata Kunci:** Analisis Kluster, Metode *Single linkage*, Metode *K-Means*, Jarak *Euclid*

## PENDAHULUAN

Indonesia merupakan negara agraris yang memiliki sumber daya alam yang beraneka ragam dan memiliki wilayah yang cukup luas. Letak astronomis negara Indonesia  $6^{\circ}$  LU –  $11^{\circ}$  LS dan  $94^{\circ}$  BT –  $141^{\circ}$  BT menandakan bahwa wilayah Indonesia merupakan wilayah yang subur dan beriklim tropis. Iklim di Indonesia yang cukup dalam memperoleh sinar matahari sepanjang tahun, mempengaruhi tumbuh suburnya setiap tanaman dengan mudah. Potensi wilayah yang demikian sangat baik kaitannya dalam pengembangan sektor pertanian (Anonim, 2009).

Sektor pertanian berperan penting pula dalam perekonomian Provinsi Bengkulu karena merupakan sektor utama yang memberikan peranan terbesar dalam pembentukan Produk Domestik Regional Bruto (PDRB). Pada tahun 2013 peranan sektor pertanian terhadap PDRB Provinsi Bengkulu adalah 39,84 persen dengan nilai nominal 8,4 triliun rupiah (atas dasar harga berlaku). Sektor pertanian merupakan kegiatan ekonomi yang menonjol untuk daerah ini bila dibandingkan dengan kegiatan di sektor lain-lainnya (Hamidi, 2013).

Salah satu kegiatan pertanian yang memiliki potensi untuk pertumbuhan baru adalah komoditas hortikultura. Hortikultura merupakan suatu cabang dari ilmu pertanian yang mempelajari budidaya buah-buahan, sayuran dan tanaman hias. Provinsi Bengkulu juga memiliki keragaman produksi tanaman hortikultura seperti sayuran dan buah-buahan. Sayuran merupakan komoditas pertanian yang memiliki harga cukup tinggi di pasaran. Tanamansayuran dapat secara nyata mendatangkan keuntungan bagi petani, sehingga banyak petani Indonesia yang membudidayakan tanaman sayuran (Hernanto, 2005).

Terdapat 8 jenis komoditi sayuran yang tercatat ditanam di berbagai daerah kabupaten/kota di Provinsi Bengkulu, yaitu bawang merah, cabe, bawang daun, tomat, terong, ketimun, kangkung dan bayam. Total luas panen tanaman sayuran pada tahun ini mencapai 997.153 hektar dengan produksi sebesar 2.251,635 ton (Hamidi, 2013). Namun, produktivitas tanaman hortikultura khususnya sayuran pada setiap kecamatan tentunya memiliki karakteristik yang berbeda-beda, melihat luas areal tanaman menghasilkan, iklim/cuaca, kesuburan tanah, dll. Melihat kondisi tersebut, maka perlu adanya pengelompokan komoditi dan kabupaten berdasarkan luas areal dan produksi yang berguna untuk meningkatkan produktivitas tanaman hortikultura di Provinsi Bengkulu (Anonim, 2009).

Analisis kluster sering kali ditemui di kehidupan sehari-hari, baik yang terkait dengan bidang sosial, bidang kesehatan, bidang *marketing*, maupun bidang pertanian. Analisis kluster merupakan suatu teknik yang dapat digunakan untuk mengklasifikasikan obyek atau kasus ke dalam kelompok yang relatif homogen. Obyek atau kasus dalam setiap kelompok cenderung mirip satu sama lain dan berbeda jauh (tidak sama) dengan obyek dari kluster lainnya. Analisis kluster juga disebut analisis klasifikasi atau taksonomi numerik (Oktarina, 2014).

Ditinjau dari hal-hal yang dikelompokkan analisis kluster dibagi menjadi dua macam yaitu pengelompokan observasi dan pengelompokan variabel. Sedangkan secara umum analisis kluster dibagi menjadi dua metode yaitu metode hierarki dan metode non hierarki. Di dalam metode hierarki sendiri terdapat beberapa metode, metode-metode yang termasuk dalam metode hierarki adalah metode Pautan Tunggal (*Single Linkage*), metode Pautan Lengkap (*Complete Linkage*), metode Antar Pusat (*Centroid Linkage*), metode Pautan Rata-rata (*Average Linkage*) dan Metode Ward (*Ward's Method*), sedangkan metode yang termasuk dalam metode non hierarki adalah metode *K-Means* (Soraya, 2011).

Metode dalam analisis kluster mempunyai kelemahan dan kelebihan sendiri - sendiri. Diantaranya adalah jika ditinjau dari proses pengelompokannya, pada metode hierarki proses pengelompokannya terjadi secara alami sehingga kelompoknya terbentuk dengan sendirinya dan digambarkan dalam bentuk *Dendogram*, sedangkan pada metode non hierarki jumlah kelompok yang terbentuk harus ditentukan terlebih dahulu oleh peneliti.

Berdasarkan uraian diatas, penulis tertarik untuk meneliti tentang Analisis Kluster dengan Menggunakan Metode *Single Linkage* dan Metode *K-Means*. Guna mengelompokkan komoditi dan kabupaten berdasarkan produksi, produktivitas, dan luas panen tanaman sayuran, agar dapat meningkatkan produktivitas tanaman hortikultura di Provinsi Bengkulu.

## Tinjauan Pustaka

### Analisis Kluster

Analisis kluster adalah suatu metode statistik yang mengidentifikasi kelompok sampel berdasarkan karakteristik serupa. Analisis kluster mengelompokkan elemen mirip sebagai objek penelitian yang mempunyai tingkat homogenitas yang tinggi antar objek menjadi kluster yang berbeda dengan tingkat heterogenitas obyek yang tinggi antar kluster. Pengklasteran ini didasarkan pada gugus variabel yang dipertimbangkan untuk diteliti (Nugroho, 2008).

Analisis kluster merupakan suatu metode yang membagi individu ke dalam kelompok yang bermakna. Analisis kluster mengelompokkan obyek berdasarkan informasi yang diperoleh pada data yang menggambarkan obyek dan keterkaitannya (Johnson and Wichern, 2002).

Analisis kluster harus memenuhi dua asumsi yaitu:

1. Sampel yang diambil harus benar-benar dapat mewakili populasi.
2. Tidak ada multikolinieritas (korelasi antar obyek), jika ada maka nilai multikolinieritasnya kurang dari 0,5 (Nugroho, 2008).

### Proses Analisis Kluster

#### 1. Standarisasi atau Transformasi Variabel

Standarisasi variabel dilakukan apabila terdapat perbedaan satuan nilai yang besar antar variabel. Misalnya ada yang dalam satuan juta dan ada yang satuan puluhan bahkan lebih kecil. Perbedaan data yang besar akan menyebabkan perhitungan jarak menjadi tidak valid sehingga data harus ditransformasi.

Transformasi dapat dilakukan dengan uji *z-score* (Nugroho, 2008), yaitu transformasi data dalam bentuk normal baku  $N(0,1)$  yang dirumuskan sebagai berikut :

$$z_i = \frac{x_i - \bar{x}}{s} \quad \text{untuk } i = 1, 2, \dots, n \quad (1.1)$$

Keterangan :

$x_i$  = data ke  $i$

$\bar{x}$  = rata - rata data

$s$  = simpangan baku

#### 2. Pengukuran Konsep Kemiripan Jarak antar Objek

Metode ini berbentuk matrix simetri  $n \times n$  yang berisi kemiripan atau ketakmiripan. Konsep ketakmiripan pada metode ini berdasarkan pada ukuran jarak antar objek, di mana jarak yang besar menunjukkan sedikit kesamaan dan jarak yang pendek atau kecil menunjukkan bahwa suatu objek semakin mirip dengan objek yang lainnya. Fungsi jarak objek  $a$  dengan objek  $b$  dapat dinotasikan dengan  $d(a, b)$  yang memenuhi sifat-sifat berikut (Nugroho, 2008).

1.  $d(a, b) \geq 0$
2.  $d(a, a) = 0$
3.  $d(a, b) = d(b, a)$
4.  $(a, b)$  meningkat seiring tidak miripnya  $a$  dan  $b$
5.  $d(a, c) \leq d(a, b) + d(b, c)$

Menurut Nugroho (2008), pengukuran jarak ada bermacam-macam, pada penelitian ini jarak yang akan digunakan adalah jarak *Euclid*.

Jarak *Euclidean* adalah besarnya jarak suatu garis lurus yang menghubungkan antar obyek. Jarak *Euclid* merupakan akar jumlah kuadrat perbedaan nilai untuk tiap variabel. Jarak *Euclid* dirumuskan sebagai berikut (Nugroho, 2008).

Jika  $\mathbf{x}' = (x_1, x_2, \dots, x_p)$  dan  $\mathbf{y}' = (y_1, y_2, \dots, y_p)$  maka  $d(x, y)$  adalah jarak antara objek  $x$  dengan objek  $y$  (Johnson and Wichern, 2002).

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$$

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})'(\mathbf{x} - \mathbf{y})} = \sqrt{\sum_{i=1}^p (x_i - y_i)^2} \quad (1.2)$$

Keterangan :

$d(\mathbf{x}, \mathbf{y})$	= jarak antara obyek $x$ dan $y$
$x_i$	= nilai obyek $x$ pada variabel $i$
$y_i$	= nilai obyek $y$ pada variabel $i$
$p$	= banyak variabel yang diamati

## Metode Pengklasteran

### 1. Metode Hierarki

Metode kluster hierarki (*hierarchical clustering methods*) adalah metode yang digunakan untuk mengelompokkan suatu pengamatan secara terstruktur berdasarkan kemiripan sifat dan banyaknya kelompok yang dapat dibentuk belum diketahui. Untuk penyajian banyaknya kelompok yang dapat dibentuk yaitu dari kelompok terbesar hingga terkecil dan sebaliknya. Pengklasteran hierarki dimulai dengan satu kluster yang diperoleh dari hasil pengamatan suatu objek dan diakhiri dengan kluster yang berisi semua objek atau sebaliknya (Rencher, 2002).

### Metode Penggabungan (*Agglomerative Methods*)

Metode penggabungan (*agglomerative methods*) merupakan metode pengklasteran yang dilakukan dari pengamatan objek-objek yang memiliki kesamaan sehingga bergabung menjadi suatu kluster kecil dan dari kluster-kluster kecil yang terbentuk akan digabungkan menjadi satu kluster besar yang berisikan semua kluster (Rencher, 2002).

### *Single Linkage (Pautan Tunggal)*

Metode pautan tunggal adalah proses pengklasteran yang didasarkan pada jarak terdekat antar objek. Jika dua objek terpisah oleh jarak yang pendek, maka kedua objek tersebut akan digabung menjadi satu kluster. Metode ini dimulai dengan menentukan objek-objek yang memiliki jarak terkecil dalam matriks *proximity*  $\mathbf{D} = \{d_{ik}\}$ . Objek-objek yang memiliki jarak terkecil digabungkan dalam satu kluster, misal  $U$  dan  $V$  adalah dua objek yang akan diklasterkan sehingga diperoleh kluster  $(UV)$ . Untuk mencari jarak antara kluster  $(UV)$  dan kluster  $W$  atau kluster lainnya diperoleh dengan rumus:

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\} \quad (1.3)$$

Keterangan:

$d_{UW}$	adalah jarak terdekat dari kluster $U$ dan $W$
$d_{VW}$	adalah jarak terdekat dari kluster $V$ dan $W$

### 2. Analisis Kluster Non Hierarki

Analisis kluster dengan metode non hierarki merupakan metode kluster yang menentukan jumlah kluster secara manual (Baroroh, 2012). Teknik analisis kluster nonhierarki dirancang untuk mengelompokkan item bukan variabel menjadi kumpulan  $K$  kluster. Banyaknya kluster  $K$ , ditentukan terlebih dahulu untuk memulai prosedur pengklasteran. Metode ini dapat diterapkan pada himpunan data yang jauh lebih besar dibandingkan menggunakan teknik analisis kluster hierarki.

Metode analisis kluster non hierarki berhubungan dengan *K-Means*, ada tiga pendekatan yang digunakan untuk menempatkan masing-masing observasi dalam satu kluster yaitu:

- a. *Sequential Threshold*
- b. *Parallel Threshold*
- c. *Optimization*

### **K – Means**

*K-Means* merupakan metode analisis kluster non hierarki yang dimulai dengan menentukan terlebih dahulu jumlah kluster yang diinginkan. Setelah jumlah kluster diketahui, baru proses kluster dilakukan tanpa mengikuti proses hierarki (Santoso, 2010). *K-Means* banyak digunakan di berbagai bidang karena sederhana dan mudah diimplementasikan. Metode ini mengklusterkan secara *partitioning* yang memisahkan data ke kelompok yang berbeda. Pembagian data biasanya dibentuk ke dalam 2 atau lebih kelompok data (Supranto, 2004).

Penggunaan kluster *K-Means* untuk menjelaskan algoritma dalam penentuan suatu obyek ke dalam kluster tertentu berdasarkan rataan terdekat. Asumsikan  $n$  adalah obyek dan  $p$  adalah variabel, dinotasikan  $x(i, j)$  adalah data obyek ke  $i$  dan variabel ke  $j$ , dengan  $i = 1, 2, \dots, n$  dan  $j = 1, 2, \dots, p$  dengan menggunakan jarak *euclid* antar objek.

Jika  $p(n, k)$  adalah partisi yang merupakan hasil pada tiap objek dialokasikan untuk salah satu dari kluster ke  $1, 2, \dots, k$ . Rata-rata variabel ke  $j$  pada kluster ke  $l$  dinotasikan dengan  $\bar{x}(j, l)$ . Maka jarak antara objek ke  $i$  dan kluster ke  $l$  didefinisikan sebagai berikut:

$$d(i, l) = \left( \sum_{j=1}^p [x(i, j) - \bar{x}(j, l)]^2 \right)^{1/2} \quad (1.4)$$

dan error partisi:

$$E[p(n, k)] = \sum_{i=1}^n D[i, l(i)]^2 \quad (1.5)$$

$E[p(n, k)]$  adalah error partisi,  $l(i)$  adalah kluster yang memuat obyek ke  $i$ ,  $D[i, l(i)]$  adalah jarak *euclid* antara obyek ke  $i$  dan rata-rata kluster yang memuat obyek (Nugroho, 2008).

### **Interpretasi Profil**

Interpretasi profil meliputi pengujian tiap kluster untuk menamai dan menandai dengan suatu label yang secara akurat dapat menjelaskan kealamian kluster. Tahap ini menggambarkan karakteristik tiap kluster untuk menjelaskan kluster-kluster tersebut dapat berbeda pada dimensi yang relevan. Membuat interpretasi dan profil kluster tidak hanya untuk memperoleh suatu gambaran saja melainkan menyediakan suatu rata-rata untuk menilai korespondensi pada kluster yang telah terbentuk dan profil kluster memberikan arahan bagi penilaian terhadap signifikansi (Nugroho, 2008).

## **METODE PENELITIAN**

### **Data**

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) yang telah dipublikasikan dalam bentuk buku dengan judul Sektor Pemerintah dalam Pertanian Indonesia. Data yang digunakan berupa data produksi tanaman sayuran, produktivitas tanaman sayuran, dan luas panen tanaman sayuran di Provinsi Bengkulu perKabupaten pada Tahun 2013.

### **Tahapan Analisis Data**

Tahapan – tahapan yang dilakukan untuk menganalisis data adalah sebagai berikut:

1. Mempartisi data untuk dikelompokan
2. Melakukan uji multikolinieritas dengan menggunakan korelasi dan *Variance Inflation Factor* (VIF)
3. Memilih ukuran jarak antar objek dengan menggunakan jarak euclid dan mengolah data menggunakan analisis kluster hierarki yaitu metode *Single Linkage* dan non hierarki yaitu metode *K-Means* dengan bantuan program R 2.14.1.
4. Melakukan interpretasi pada kluster yang telah di peroleh.

### **Hasil dan Pembahasan**

#### **Deskripsi Data**

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) di Provinsi Bengkulu Tahun 2013. Data yang digunakan terdiri dari 3 data yang berbeda yaitu, data produksi, produktivitas, dan luas panen tanaman sayuran. Variabel yang digunakan dalam penelitian ini ada 8 variabel yang terdiri dari, bawang merah, cabe, bawang daun, tomat, terung, ketimun,

kangkung, dan bayam. Sedangkan obyek-obyek yang digunakan adalah 10 Kabupaten yang berada di Provinsi Bengkulu.

### Uji Multikolinieritas

Multikolinieritas terjadi bila terdapat dua atau lebih variabel yang saling berkaitan. Untuk mengetahui ada tidaknya multikolinieritas antar variabel dapat dilakukan dengan matrik korelasi. Pada penelitian ini untuk mengetahui ada tidaknya multikolinieritas, penulis menggunakan uji korelasi dan uji *Variance Inflation Factor* (VIF) untuk mendeteksi masalah multikolinieritas. Jika nilai korelasi mendekati satu atau lebih besar dari 0,5 maka dapat dikatakan terdapat masalah multikolinieritas.

Nilai matrix korelasi dari output program R dapat dilihat pada lampiran 3, dari matrix korelasi pada lampiran 3 terdapat nilai korelasi yang lebih dari 0,5 dan terdapat matrix korelasi yang nilainya kurang dari 0,5. Variabel yang memiliki nilai matrix korelasi kurang dari 0,5 dapat dilihat pada tabel 1 berikut:

Tabel 4 Matrix Korelasi Data Produksi Tanaman Sayuran Tahun 2013

	Cabe	Kangkung	Ketimun	Terung	Tomat
B. Daun	0,0974736	0,396984	0,41681823	0,0238983	
B. Merah	0,007417523	0,313822489	0,321295434	0,081170266	
Bayam	0,1765056		0,4599876	0,1118327	
Cabe					0,2226143
Terung					0,14424849

Berdasarkan data produksi tanaman sayuran maka didapatkan nilai *Variance Inflation Factor* (VIF) yang diperoleh dari output program R dan mendapatkan hasil sebagai berikut.

Tabel 2 Nilai *Variance Inflation Factor* (VIF) Data Produksi Tanaman Sayuran dari Output Program R

No	Variabel	VIF
1	Bawang Merah	6,241815
2	Ketimun	1,697112
3	Kangkung	2,376218
4	Bayam	8,140464

### Metode Pengklasteran

Pengklasteran pada penelitian ini menggunakan metode Hierarki dan metode Non Hierarki, pada hierarki metode yang digunakan yaitu metode *Single Linkage*, sedangkan pada non hierarki metode yang digunakan yaitu *K-Means*, jarak yang digunakan oleh kedua metode ini adalah Jarak *Euclid*.

### Metode Hierarki (*Single Linkage*)

Pengklasteran pada penelitian ini menggunakan metode Hierarki yaitu Metode *SingleLinkage* dan jarak yang digunakan adalah Jarak *Euclid*, tabel dibawah ini menunjukkan masing-masing nilai dari jarak *Euclid* yang dihasilkan oleh 3 data.

Tabel 3 Nilai Jarak *Euclid*, pada Output Dari Program R

	Produksi	Produktivitas	Luas Panen
<b>Min</b>	2145,4	236,63	48,888
<b>1st Qu</b>	4930,8	614,02	256,280
<b>Median</b>	6745,6	937,31	369,320
<b>Mean</b>	59197,0	1299,90	756,790
<b>3rd Qu</b>	73069,0	1746,00	536,030
<b>Max</b>	216430,0	3240,60	2586,800

Berdasarkan Tabel 11 dapat dilihat nilai-nilai dari masing-masing jarak *Euclid* yang dihasilkan oleh 3 data. Dari jarak tersebut dapat dilihat bahwa terdapat perbedaan dari nilai minimum, kuartil pertama,

*median, mean, kuartil ketiga dan nilai maksimum pada masing-masing data yang dihasilkan oleh jarak Euclid.*

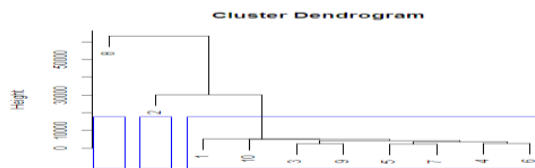
**Metode *Single Linkage* untuk Data Produksi Tanaman Sayuran**

Tabel 4 Proses Agglomerative metode *Single Linkage* dan nilai Jarak *Euclid* pada Output Dari Program R

Step	Clusters	Joined	Order of Object	Distance Level
1	5	7	3	5171,139
2	3	9	9	2462,161
3	4	6	4	4211,873
4	3	1	6	2662,936
5	2	4	5	3498,428
6	5	10	7	2145,444
7	1	6	10	4894,157
8	7	8	8	68470,311
9	8	2	2	204976,039

Setelah proses *Agglomerative* terbentuk maka pengklasteran akan dilakukan dengan menggabungkan semua kluster-kluster kecil yang sudah terbentuk menjadi satu kluster besar yang berisikan semua kluster. Jika kluster besar sudah terbentuk maka hasil pengklasteran akan muncul dalam bentuk dendrogram.

Dendrogram merupakan visualisasi proses pengklasteran yang terjadi. Posisi garis pada skala dendrogram menunjukkan jarak pada kluster yang sudah tergabung. Dendrogram yang diperoleh pada penelitian ini adalah sebagai berikut:



Gambar 1. Dendrogram Metode *Single Linkage* menggunakan Jarak *Euclid*

Pada dendrogram metode *Single Linkage* tahap pertama objek (Bengkulu Utara dan Bengkulu Tengah), objek (Seluma dan Lebong), objek (Kaur dan Muko-Muko), membentuk satu kluster tersendiri, karena mempunyai panjang garis yang sama dan tergabung menjadi satu kesatuan. Demikian seterusnya sampai akhirnya objek akan bergabung menjadi beberapa kluster.

Pada metode *Single Linkage*, kluster yang dapat dibentuk adalah 1,2 dan 3 kluster, dimana jika kluster 1 yang diambil maka anggotanya adalah kluster dengan objek (Bengkulu Selatan), objek (Kota Bengkulu), objek (Bengkulu Utara dan Bengkulu Tengah), objek (Seluma dan Lebong) dan objek (Kaur dan Muko-Muko). Jika kluster 2 yang diambil maka anggota kluster tersebut adalah objek (Rejang Lebong) dan 9 objek lainnya. Jika kluster 3 yang diambil maka anggotanya adalah kluster dengan objek (Kepahiang), dan 8 objek lainnya. Pada metode *Single Linkage* ini, Rejang Lebong membentuk kluster tunggal dengan anggota dirinya sendiri, dan Kepahiang juga membentuk kluster tunggal dengan anggota dirinya sendiri. Hal ini berarti objek Rejang Lebong dan Kepahiang berbeda jauh dengan objek pada kluster lainnya. Hal ini dikarenakan Rejang Lebong memiliki Hasil Produksi cabe dan daun bawang terbesar dari Provinsi lainnya. Dan objek Kepahiang juga memiliki Hasil Produksi cabe terbesar dari Provinsi lain.

**Metode Non Hierarki (*K-Means*)**

Metode *K-Means* memproses semua data secara sekaligus dimana *k* adalah banyaknya kluster. Dengan menggunakan R 2.14.1 maka data akan diproses.

**1. Metode *K-Means* untuk Data Produksi Tanaman Sayuran**

Berikut proses yang dilakukan dalam analisis klster metode *K-means* meliputi:

- a. Menentukan nilai *K* sebagai jumlah klsater yang akan dibentuk

Dari output program R untuk proses pengklasteran dengan menggunakan metode *K-means* kluster yang akan dibentuk yaitu 3 kluster dengan memperhatikan semua kemungkinan pada data.

b. Menentukan nilai *centroid* awal secara *random*

Berdasarkan kluster yang akan dibentuk (*K*) pada proses pengklasteran metode *k-means* yang diambil secara acak. Sehingga terdapatlah nilai *centroid* awal metode *k-means* yang diperoleh/ditentukan dari tiga objek pengamatan yaitu, *centroid* pertama berupa Kabupaten Kepahiang, *centroid* kedua Kabupaten Rejang Lebong, *centroid* ketiga Kabupaten Kaur. Sehingga didapat nilai *centroid* awal sebagai berikut.

Tabel 5 Nilai *Centroid* awal metode *K-Means*

Variabel	Centroid Kluster		
	1	2	3
X <sub>1</sub>	0	692	0
X <sub>2</sub>	54646	9565	704
X <sub>3</sub>	20206	215140	305
X <sub>4</sub>	4554	16342	116
X <sub>5</sub>	44636	4136	1306
X <sub>6</sub>	20608	10325	175
X <sub>7</sub>	4756	3597	2572
X <sub>8</sub>	2306	7100	1213

c. Menentukan jarak setiap objek dengan setiap *centroid* pada metode *K-means*

Untuk menghitung jarak setiap objek dengan *centroid* pertama, kedua, dan ketiga, kita dapat menggunakan rumus jarak *euclid* pada persamaan (2.2). Adapun contoh perhitungan jarak pada setiap objek dengan *centroid* adalah sebagai berikut:

- Kabupaten Bengkulu Selatan dengan *Centroid* pertama

$$= \sqrt{(0 - 0)^2 + (4695 - 54646)^2 + (122 - 20206)^2 + \dots + (1661 - 2306)^2}$$

$$= 68467,27247$$

- Kabupaten Rejang Lebong dengan *centroid* pertama

$$= \sqrt{(692 - 0)^2 + (9568 - 54646)^2 + (21514 - 20206)^2 + \dots + (7100 - 2306)^2}$$

$$= 63205,30863$$

Lakukan cara yang sama untuk perhitungan jarak setiap objek dengan *centroid* pertama, kedua, dan ketiga, sampai objek ke 10. Sehingga diperoleh keseluruhan nilai jarak *euclid* dari masing-masing objek pada setiap *centroid* adalah sebagai berikut.

Tabel 6 Nilai jarak *Euclid* metode *K-means*

Objek	Kluster		
	1	2	3
Bengkulu Selatan	68467,27247	76272,7633	8241,167939
Rejang Lebong	63205,30863	0	31744,9797
Bengkulu Utara	71396,87443	75819,70775	5986,714124
Kaur	75009,63873	75989,5278	0
Seluma	72904,48184	76521,89155	3408,787321
Muko - Muko	74154,92218	75735,03059	2641,522856
Lebong	73032,81691	76880,78511	4755,261192
Kepahiang	0	92994,86998	75007,15603
Bengkulu Tengah	72563,90395	74606,1622	4527,419795
Kota Bengkulu	69770,10153	74621,09128	6138,120071

d. Mengelompokan objek berdasarkan jarak minimum

Misalkan angka 1 untuk nilai jarak yang kecil, sedangkan 0 untuk nilai jarak yang yang besar, maka dapat dilihat pada matriks berikut.

$$G = \begin{pmatrix} A & B & C & D & E & F & G & H & I & J \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \end{pmatrix}$$

Berdasarkan uraian diatas dapat dilihat anggota masing-masing kluster yang terbentuk, kluster pertama beranggotakan Kabupaten Kepahiang, kluster dua beranggotakan Kabupaten Rejang Lebong, sedangkan kluster ketiga beranggotakan delapan Kabupaten lainnya.

e. Iterasi 1, menentukan *centroid* baru

Berdasarkan hasil kluster yang terbentuk diatas, maka kita akan menentukan *centroid* baru, dengan cara yang sama pada langkah (b), maka diperoleh nilai *centroid* baru sebagai berikut.

Tabel 7 Nilai *Centroid* metode *K-Means*

Variabel	Centroid Kluster		
	1	2	3
X <sub>1</sub>	0	692	7
X <sub>2</sub>	54646	9565	3667
X <sub>3</sub>	20206	215140	90
X <sub>4</sub>	4554	16342	655
X <sub>5</sub>	44636	4136	2362
X <sub>6</sub>	20608	10325	1081
X <sub>7</sub>	4756	3597	1544
X <sub>8</sub>	2306	7100	1323

f. Iterasi 1, Menghitung jarak setiap objek dengan setiap *centroid* pada metode *K-means*

Lakukan cara yang sama pada langkah (c) untuk menghitung jarak setiap objek dengan *centroid* pertama, kedua, dan ketiga, kita dapat menggunakan rumus jarak *euclid* pada persamaan (2.2). Adapun contoh perhitungan jarak pada setiap objek dengan *centroid* adalah sebagai berikut:

- Kabupaten Bengkulu Selatan dengan *Centroid* pertama

$$= \sqrt{(0 - 0)^2 + (4695 - 54646)^2 + (122 - 20206)^2 + \dots + (1661 - 2306)^2}$$

$$= 68467,27247$$

- Kabupaten Rejang Lebong dengan *centroid* pertama

$$= \sqrt{(692 - 0)^2 + (9568 - 54646)^2 + (21514 - 20206)^2 + \dots + (7100 - 2306)^2}$$

$$= 63205,30863$$

Lakukan cara yang sama untuk perhitungan jarak setiap objek dengan *centroid* pertama, kedua, dan ketiga, sampai objek ke 10. Sehingga diperoleh keseluruhan nilai jarak *euclid* dari masing-masing objek pada setiap *centroid* adalah sebagai berikut.

Tabel 8 Nilai jarak *Euclid* metode *K-means*

Objek	Kluster		
	1	2	3
Bengkulu Selatan	68467,27247	76272,7633	5694,124779
Rejang Lebong	63205,30863	0	30549,31299
Bengkulu Utara	71396,87443	75819,70775	3899,311991
Kaur	75009,63873	75989,5278	3481,525528
Seluma	72904,48184	76521,89155	1632,812604
Muko - Muko	74154,92218	75735,03059	3389,776836
Lebong	73032,81691	76880,78511	3058,118866
Kepahiang	0	92994,86998	72036,22701
Bengkulu Tengah	72563,90395	74606,1622	2983,102412
Kota Bengkulu	69770,10153	74621,09128	3740,406128

g. Iterasi 1, Mengelompokan objek berdasarkan jarak minimum

Misalkan angka 1 untuk nilai jarak yang kecil, sedangkan 0 untuk nilai jarak yang yang besar, maka dapat dilihat pada matriks berikut.

$$G = \begin{matrix} & \begin{matrix} A & B & C & D & E & F & G & H & I & J \end{matrix} \\ \begin{matrix} A \\ B \\ C \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \end{pmatrix} \end{matrix}$$

Berdasarkan uraian diatas dapat dilihat anggota masing-masing kluster yang terbentuk, kluster pertama beranggotakan Kabupaten Kepahiang, kluster dua beranggotakan Kabupaten Rejang Lebong, sedangkan kluster ketiga beranggotakan delapan Kabupaten lainnya.

Setelah melakukan iterasi pertama ternyata tidak terdapat perpindahan dari masing-masing anggota kluster yang terbentuk, karena tidak terdapat perpindahan maka proses kluster dihentikan, untuk melihat hasil perhitungan pada output program R dapat dilihat pada lampiran 8 (a).



## Interpretasi

Interpretasi profil klaster adalah proses pemberian nama dari masing-masing klaster yang telah terbentuk. Nama yang diberikan sesuai keperluan peneliti, dengan syarat bahwa nama yang diberikan kepada klaster tersebut mewakili karakteristik atau ciri dari kelompok (Supranto, 2010).

## Kesimpulan

Dari hasil penelitian yang dilakukan dengan judul Analisis Klaster Dengan Menggunakan Metode *Single Linkage* dan Metode *K-Means* dengan menggunakan jarak *Euclid*, pengklasteran yang dilakukan dengan menggunakan tiga data yang berbeda yaitu Produksi, Produktivitas, dan Luas Panen Tanaman Sayuran ini menghasilkan pemecahan klaster yang berbeda antara metode *Single Linkage* dan metode *K-means*. Sehingga dapat disimpulkan sebagai berikut:

1. Dari hasil pengklasteran yang terbentuk dengan menggunakan metode *Single Linkage* dan metode *K-means*, berdasarkan Data Produksi, Produktivitas, dan Luas Panen Tanaman Sayuran, masing-masing metode membentuk 3 klaster yang berbeda dari setiap data tersebut. Pada metode *Single Linkage* klaster yang terbentuk tidak mendapatkan perbedaan yang jauh antara setiap data yang digunakan. Sedangkan metode *K-means* klaster yang terbentuk mendapatkan perbedaan yang jauh antar setiap data yang digunakan.
2. Pada penelitian ini dapat dijelaskan bahwa metode yang baik dalam analisis klaster yang terbentuk adalah metode *Single Linkage*, karena metode *Single Linkage* tidak mendapatkan perbedaan yang jauh antar setiap klaster yang terbentuk. Sedangkan Kabupaten yang paling berpengaruh pada penelitian ini adalah Kabupaten Rejang Lebong, karena selalu menghasilkan hasil yang baik dalam setiap proses klaster yang telah dibentuk.

## Saran

Pada tulisan ini penulis menggunakan metode *Single Linkage* dan *K-Means* dalam bidang ilmu pertanian. Penelitian selanjutnya dapat juga menggunakan data dalam bidang ilmu lain dengan menggunakan perbandingan metode *Single Linkage* dan *K-Means* dengan metode pengklasteran lain mengingat cakupan klaster yang sangat luas.

## DAFTAR PUSTAKA

- Anonim. 2009. *Pertanian Indonesia*. Penebar Swadaya. Jakarta.
- Hamidi, H. A. 2013. *Jurnal Optimis Pola Tanaman Usaha Tani Sayuran Selada dan Sawi di Daerah Produksi Padi*. Fakultas Pertanian Universitas Syiah Kuala. Banda Aceh.
- Hernanto, F. 2005. *Ilmu Usaha Tani*. Penebar Swadaya. Jakarta.  
<https://Pengertian.Produktivitas.kerja.htm>(Diakses 01 februari 2015).
- Johnson, R. A.,and D, W. Wichern. 2002. *Applied Multivariate Statistical Analysis, Fifth Edition*. Pearson Education. United States Of America.
- Nugroho, S. 2008. *Statistika Multivariat terapan, edisi pertama*. UNIB Press, Universitas Bengkulu.
- Oktarina, 2014. *Metode Linkage pada Persentase Kelahiran Balita Menurut Penolong Kelahirn Terakhir*. Skripsi, Program Studi Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam. Universitas Bengkulu.
- Rencher, A. C. 2002. *Methods Of Multivariate Analysis Second Edition*. Wiley Series In Probability and Mathematical Statistics: Canada.
- Soraya, Y. 2011. *Perbandingan Kinerja Metode Single Linkage Metode Complate Linkage Metode K-Means Dalam Analisis Klaster*. Jurusan Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang.
- Supranto, J. 2004. *Analisis Multivariat Arti dan Interpretasi*. Rineka Cipta: Jakarta.