

M-14

## ANALISIS K-MEANS CLUSTER UNTUK PENGELOMPOKAN KABUPATEN /KOTA DI JAWABARAT BERDASARKAN INDIKATOR MASYARAKAT

Soemartini<sup>1</sup>, dan Enny Supartini<sup>2</sup>

<sup>1,2)</sup> Departemen Statistika FMIPA UNPAD Bandung  
[tine\\_soemartini@yahoo.com](mailto:tine_soemartini@yahoo.com), [arthinii@yahoo.com](mailto:arthinii@yahoo.com)

### Abstrak

*Analisis kluster merupakan salah satu metode Multivariat yang bertujuan untuk mengelompokkan objek berdasarkan kemiripan atau tidaknya berdasarkan karakteristiknya. Penelitian ini bertujuan untuk mengelompokkan 27 Kab/Kota di Jawa Barat berdasarkan indikator kesejahteraan Rakyat dan untuk mendapatkan prioritas dalam program-program pembangunan dari pemerintah agar tepat sasaran, yang sangat tergantung pada ketepatan pengidentifikasian target group dan target area . Dengan menggunakan K-Means yang merupakan salah satu metode pengklasteran yang paling banyak digunakan sebagai alternative metoda klaster, Hal ini dikarenakan memiliki ketepatan yang lebih tinggi jika dibandingkan dengan metode Hirarki. Metoda K –Means dapat digunakan untuk menjelaskan algorima dalam penentuan suatu objek ke dalam kluster tertentu berdasarkan rataan terdekat dan mudah diimplementasikan dan berdasarkan hasil, kluster pertama terdiri 8 Kab/kota, dan kluster kedua terdiri dari 19 Kab/kota dengan nilai CCR sebesar 96.3 % dan APER sebesar 3,7 % ini menunjukkan bahwa klasifikasi yang terbentuk memiliki tingkat ketepatan yang sangat tinggi juga merupakan pengklasteran secara partitioning yang memisahkan ke dalam kelompok yang berbeda.*

*Kata kunci : APER ,CCR dan K-Means Cluster*

### 1. PENDAHULUAN

Kesejahteraan merupakan tumpuan harapan dan menjadi cita-cita luhur perjuangan bangsa Indonesia ,selain itu kesejahteraan merupakan hal yang menentukan suatu pembangunan di suatu daerah. Kesejahteraan masyarakat diharapkan meningkat dari tahun ke tahun. Kesejahteraan mengandung pengertian yang relatif, dinamis, dan kuantitatif. Rumusannya tidak pernah final karena akan terus berkembang seiring dengan perkembangan kebutuhan hidup manusia. Secara umum kesejahteraan dapat diartikan sebagai suatu keadaan dimana segenap warga negara selalu berada dalam kondisi serba kecukupan segala kebutuhannya, baik material maupun spiritual ( Faisal Basri ,1993) .

Seperti yang kita ketahui Jawa Barat merupakan salah satu propinsi di Indonesia yang penduduk miskinnya pada bulan September 2014 memiliki persentase 9,18 persen.mengalami penurunan dibandingkan pada bulan Maret 2014 sebesar 9,44 persen.

Meskipun jumlah penduduk miskin menurun, pada kenyataannya pertumbuhan kesempatan kerja tidak mampu menyerap pertumbuhan angkatan kerja dalam setiap tahunnya. Kemiskinan merupakan topik strategis dan

mendapatkan prioritas utama untuk ditangani. Disamping masalah kemiskinan, angka pengangguran di Jawa Barat juga masih cukup tinggi. Sesuai data tahun 2014, angka pengangguran mencapai 1.845.591 jiwa. Implikasinya, tingkat pengangguran terbuka akan semakin meningkat, jika tidak ada perubahan strategi dalam penciptaan lapangan kerja. Dalam melaksanakan program pembangunan perlu adanya identifikasi berdasarkan karakteristik tingkat kesejahteraan rakyat tiap daerah agar dalam mengambil kebijakan dan strategi pembangunan bisa tepat sasaran dan tepat guna.

Salah satu prasyarat (Faisal Basri, 1995), keberhasilan program-program pembangunan sangat tergantung pada ketepatan pengidentifikasian *target group* dan *target area*. Oleh karena itu, sangat penting mempertimbangkan pengelompokan dan karakteristik 27 kabupaten/kota di Provinsi Jawa Barat berdasarkan indikator kesejahteraan rakyat Tahun 2014.

Dalam statistika, salah satu metode yang digunakan untuk mengelompokkan variabel atau objek adalah analisis kluster. Analisis kluster merupakan suatu metode untuk mengelompokkan variabel atau objek ke dalam beberapa kelompok. Analisis kluster merupakan analisis yang digunakan untuk mengelompokkan pengamatan atau variabel menjadi beberapa kelompok pengamatan atau variabel yang jumlahnya lebih sedikit. Analisis kluster dilakukan jika ingin mengetahui jumlah kelompok baru. Analisis kluster bertujuan untuk mengelompokkan  $n$  objek berdasarkan  $p$  variat yang memiliki kesamaan karakteristik diantara objek-objek tersebut. Objek tersebut akan diklasifikasikan ke dalam satu atau lebih kluster (kelompok) sehingga objek-objek yang berada dalam satu kluster akan mempunyai kemiripan atau kesamaan karakter.

Dengan menggunakan analisis cluster K-Means dapat dilakukan pengelompokan Kabupaten/Kota berdasarkan indikator Kesejahteraan dan juga untuk mengetahui daerah mana saja di Jawa Barat yang diprioritaskan untuk mendapatkan bantuan agar tercapainya program-program pembangunan dari Pemerintah.

## 2. METODE PENELITIAN

### a. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder Bidang Statistik Sosial (BPS Prop. JABAR, 2014) tentang Indikator Kesejahteraan Rakyat Provinsi Jawa Barat 2014. Data yang digunakan berupa indikator kesejahteraan untuk kabupaten/kota di Jawa Barat, yang terdiri dari 27 kabupaten/kota.

### b. Variabel Penelitian

Variabel-variabel yang digunakan pada penelitian ini adalah sebagai berikut :

- 1) Variabel Kepadatan Penduduk (KP)

Salah satu persoalan yang terkait dengan kependudukan yang masih harus dihadapi di Jawa Barat yaitu masalah ketimpangan distribusi penduduk. Distribusi penduduk yang tidak merata menimbulkan masalah pada kepadatan penduduk dan tekanan disuatu wilayah. Hal ini sangat berpengaruh pada kondisi masyarakat setempat.

2) Variabel Angkatan Kerja (AK)

Angkatan kerja adalah tenaga kerja atau penduduk dalam usia yang berkerja, atau yang mempunyai pekerjaan namun untuk sementara sedang tidak bekerja dan mencari pekerjaan.( Dumairy,1996).

3) Variabel Laju Pertumbuhan Penduduk (LPP)

Salah satu masalah klasik kependudukan yang terjadi di Jawa Barat dan terus menjadi perhatian pemerintah adalah jumlah penduduk Jawa Barat yang besar.

4) Variabel Rata-rata Pengeluaran Perkapita (PP)

Rata-rata pengeluaran per kapita merupakan salah satu indikator penting dalam menentukan kesejahteraan penduduk pada suatu wilayah. Perubahan pendapatan seseorang akan berpengaruh pada pergeseran pola pengeluaran. Dengan demikian pola pengeluaran dapat digunakan sebagai salah satu alat untuk mengukur tingkat kesejahteraan penduduk, dimana perubahan komposisinya digunakan sebagai petunjuk perubahan tingkat kesejahteraan.

5) Variabel Angka Harapan Hidup (AHH)

Angka harapan hidup adalah rata-rata tahun hidup yang masih akan dijalani oleh seseorang yang telah berhasil mencapai umur X, pada suatu tahun tertentu, dalam situasi mortalitas yang berlaku di lingkungan masyarakatnya.

6) Variabel Rata-rata Lama Sekolah (RLS)

Rata-rata lama sekolah (RLS) adalah rata-rata jumlah tahun yang dihabiskan oleh penduduk berusia 15 tahun keatas untuk menempuh semua jenis pendidikan formal yang pernah dijalani. Indikator RLS ini dihitung dari variable pendidikan tertinggi yang ditamatkan dan tingkat pendidikan yang sedang dijalankan. Standar UNDP (Badan Program Pembangunan PBB).

**c. Metode Analisis Data**

Penelitian ini menggunakan metode analisis kuantitatif dengan analisis kluster *k-means* kemudian dilanjutkan dengan analisis diskriminan untuk mengevaluasi keakuratan klasifikasi. Sebelum melakukan kedua uji tersebut, ada beberapa asumsi yang harus terpenuhi dalam analisis cluster *k-means* dan analisis diskriminan.

**d. Analisis Klaster**

Analisis klaster merupakan suatu metode statistik yang digunakan untuk mengelompokkan sekumpulan objek kedalam kelompok-kelompok berdasarkan karakteristik yang dimilikinya. (Johnson, 1998), Objek diklasifikasikan kedalam satu atau lebih klaster sehingga objek-objek yang berada di dalam klaster mempunyai kemiripan atau kesamaan karakter .

Apabila terdapat n objek dan p variabel, maka observasi dengan  $i = 1, 2, \dots, n$  dan  $j = 1, 2, \dots, p$ , dapat digambarkan sebagai berikut:

**Tabel 2.1 Susunan Observasi Analisis Klaster**

	Variabel 1	Variabel 2	Variabel 3	...	Variabel p
Objek 1	$X_{11}$	$X_{12}$	$X_{13}$	...	$X_{1p}$
Objek 2	$X_{21}$	$X_{22}$	$X_{23}$	...	$X_{2p}$
Objek 3	$X_{31}$	$X_{32}$	$X_{33}$	...	$X_{3p}$
.	.	.	.	.	.
Objek n	$X_{n1}$	$X_{n2}$	$X_{n3}$	...	$X_{np}$

Perbedaan analisis klaster dengan analisis faktor terletak pada karakteristik objek yang disederhanakan. Analisis faktor bertujuan untuk menyederhanakan variabel atau faktor sedangkan analisis klaster bertujuan untuk menyederhanakan atau membentuk setiap kelompok berdasarkan kesamaan karakteristik observasi atau kasus ( Supranto, 2004).

**e. Memilih Ukuran Kemiripan**

Untuk mengetahui seberapa mirip objek-objek yang diteliti dibutuhkan ukuran jarak (distance-type measure). Dengan memiliki sebuah ukuran kuantitatif, maka proses pengklasteran akan lebih mudah. Pada umumnya, jarak yang biasa digunakan adalah Jarak Euclidean( Dillon,1984)

Jarak *eulidean* dirumuskan sebagai berikut :

$$d_{(i,j)} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \tag{1}$$

dengan  $i = 1, 2, \dots, 27$  dan  $j = 1, 2, 3, 4$  ;  $p$  : banyak variable yang diamati

$d_{(i,j)}$  : jarak antara objek  $i$  dan objek  $j$

$x_{ik}$  : nilai objek  $i$  pada variable ke- $k$

$x_{jk}$  : nilai objek  $j$  pada variable ke- $k$

**f. Metode K-Means Cluster**

Metode *K-means* digunakan sebagai alternative metode klaster untuk data dengan ukuran yang lebih besar. Hal ini dikarenakan metode ini memiliki kecepatan yang lebih tinggi jika dibandingkan dengan metode hirarki. Metode *K-means* dapat digunakan untuk menjelaskan algoritma dalam penentuan suatu objek kedalam klaster tertentu berdasarkan rata-rata terdekat ( Dillon , 1984).

*K-means* merupakan metode pengelompokan yang paling banyak digunakan diberbagai bidang karena sederhana dan mudah diimplementasikan. *K-means* merupakan metode pengklasteran secara *partitioning* yang memisahkan data kedalam kelompok yang berbeda. Tujuan dari pengklasteran data ini adalah untuk meminimalisasikan fungsi objektif dalam proses pengelompokkan, yang pada umumnya berusaha meminimalisasikan ragam didalam suatu kelompok dan memaksimalkan ragam antar kelompok (Febriyana 2011).

**g. Analisis diskriminan**

Analisis diskriminan merupakan suatu analisis dengan tujuan membentuk sejumlah fungsi melalui kombinasi linear variabel-variabel asal, yang dapat digunakan sebagai cara terbaik untuk memisahkan kelompok-kelompok individu. Fungsi yang terbentuk melalui analisis ini selanjutnya dinamakan fungsi diskriminan.

Analisis diskriminan juga dapat digunakan jika variabel terikat terdiri dari dua kelompok. Apabila klasifikasi terdiri dari tiga kelompok atau lebih maka teknik yang digunakan adalah analisis diskriminan multipel (*multiple discriminant analysis*). Adapun tujuan analisis diskriminan yakni, menentukan prediktor mana yang memberikan kontribusi terbesar terhadap terjadinya perbedaan antar kelompok dan mengevaluasi keakuratan klasifikasi (Johnson, 1998).

Dalam penelitian ini analisis diskriminan digunakan untuk mengetahui ketepatan klasifikasi. Jika nilai ketepatan klasifikasi besar, maka pengklasteran semakin baik. Sedangkan jika nilai ketepatan klasifikasi kecil, maka pengklasteran kurang baik.

**Tabel 2.2 Tabel Klasifikasi**

D (Diskriminan)		Prediksi		Total
		1	2	
Aktual	1	C11	C12	C11+C12
	2	C21	C22	C21+C22

Menghitung seberapa besar klasifikasi dengan *Correct Classification Rate (CCR)*. CCR merupakan persentase ketepatan nilai pengamatan dan dugaannya, CCR dihitung dengan persamaan berikut :

$$CCR = \frac{\text{Jumlah prediksi yang tepat}}{\text{jumlah data}} \times 100\%$$

$$CCR = \frac{c_{1,1}+c_{2,2}}{\sum_{i=1}^2 \sum_{j=1}^2 c_{ij}} \times 100\% \tag{3}$$

Menghitung persentase kesalahan dalam klasifikasi dihitung menggunakan APER yaitu sebagai berikut:

$$APER = \frac{c_{1,2}+c_{2,1}}{\sum_{i=1}^2 \sum_{j=1}^2 c_{ij}} \times 100\% \tag{4}$$

Semakin kecil nilai APER maka tingkat ketepatan klasifikasi semakin baik

**h. Beberapa asumsi yang harus dilakukan dalam pengujian Cluster K-Means**

**1) Uji Multikolenieritas**

Multikolenieritas merupakan adanya hubungan yang linier diantara variabel penelitian.

$$VIF_i = \frac{i}{1-R_i^2} \tag{5}$$

Jika nilai VIF dari suatu variable memiliki nilai lebih dari 10, maka variabel tersebut mengindikasikan terjadinya multikolenieritas.

**2) Asumsi Klasik Untuk Pengujian Diskriminan**

**a) Uji Normal Multivariate**

Pengujian asumsi normal *multivariate* dilakukan pada masing-masing data tiap klaster untuk mengetahui apakah data pada setiap klaster tersebut menyebar mengikuti sebaran normal multivariate, dengan

Menghitung jarak mahalanobis ( $d_i^2$ ) pada kelompok ke-i dengan persamaan :

$$d_i^2 = (x_i - \bar{x}_i)' S^{-1} (x_i - \bar{x}_i) \tag{6}$$

Dengan  $i = 1, 2, \dots, n$

Apabila plot antara jarak Mahalanobis dan Chi-Kuadrat mengikuti pola garis lurus maka dapat dikatakan bahwa data berdistribusi *normal multivariate*.

**b) Uji Kehomogenan Matriks Kovarian Kedua Kelompok**

Pengujian kehomogenan matriks kovarian gabungan kedua kelompok diperlukan untuk dapat membentuk fungsi diskriminan. Pengujian ini dilakukan dengan Statistik uji Box'M ( $MC^{-1}$ ), dibandingkan dengan  $\chi^2_{\alpha; db}$ .

Statistik uji Box'M ( $MC^{-1}$ ), dibandingkan dengan  $\chi^2_{\alpha; db}$

$$M = \sum_{q=1}^k (n_q - 1) \ln|S| - \sum_{q=1}^k (n_q - 1) \ln|S_q| \tag{7}$$

$$C^{-1} = 1 - \frac{2p^2 + 3p - 1}{6(p+1)(k-1)} \left( \sum_{q=1}^k \frac{1}{(n_q - 1)} - \frac{1}{\sum_{q=1}^k (n_q - 1)} \right) \tag{8}$$

**c) Uji Perbedaan Vektor Rata-rata**

Uji beda vektor rata-rata kedua kelompok, yaitu rata-rata antar kedua kelompok. Pada kasus disini yaitu kelompok Kab/Kota yang sejahtera ( $\mu_1$ ) dan rata-rata kelompok Kab/Kota yang tidak sejahtera ( $\mu_2$ ), pada  $\alpha = 5\%$ .

$$\text{Statistik Uji : } T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)' S^{-1} (\bar{x}_1 - \bar{x}_2) \tag{9}$$

$$\text{Kriteria Penolakan : } \frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2 > F_{\alpha; (p, n_1 + n_2 - p - 1)} \tag{10}$$

**d) Uji Variance**

Uji *variance* dilakukan untuk melihat apakah variable-variabel yang telah membentuk klaster memiliki perbedaan pada tiap klaster, serta

untuk melihat variable mana yang paling berpengaruh terhadap pembentukan kluster

Hal ini dapat dilihat dengan menggunakan konsep sebagai berikut :

$$\bar{y}_i = \frac{y_i}{M} \quad i = 1, 2, \dots$$

$\bar{y}_i$ : rata-rata setiap variabel pada kluster ke-i

$M$  : jumlah anggota pada kluster ke-i

$$\bar{y} = \sum_{i=1}^N \sum_{j=1}^M \frac{y_{ij}}{N} \quad (11)$$

$\bar{y}$ : rata-rata populasi dari variabel

$N$  : banyaknya kluster  $\bar{y} = \frac{y_{jk}}{N}$

$\bar{y}$ : rata-rata populasi

$y_{jk}$  : nilai data ke-j pada variabel ke-k

$$\sigma_i^2 = \sum_{j=1}^M \frac{(y_{ij} - \bar{y}_i)^2}{M-1} \quad i = 1, 2 \quad (12)$$

$\sigma_i^2$ : variansi pada kluster ke-i

$y_{ij}$  : nilai data ke-j pada kluster ke-i

*Internal homogeneity*  $\rightarrow$  variance within cluster  $\sigma_w^2$

$$\sigma_w^2 = \sum_{i=1}^N \frac{\sigma_i^2}{N} \quad (13)$$

$\sigma_w^2$ : variansi dalam kluster

*External homogeneity*  $\rightarrow$  variance between cluster  $\sigma_b^2$

$$\sigma_b^2 = \sum_{i=1}^N \frac{(\bar{y}_i - \bar{y})^2}{N-1} \quad (14)$$

$\sigma_b^2$ : variansi antar kluster  $\sigma^2 = \frac{\sigma_w^2}{\sigma_b^2} \quad (15)$

Semakin besar nilai  $\sigma^2$  suatu variabel, maka semakin besar perbedaan variabel tersebut pada kluster yang terbentuk. Hal ini dapat digunakan sebagai metode pembandingan untuk mengetahui metode pengelompokan mana yang lebih baik. Semakin besar nilai  $\sigma^2$  pada variabel semakin baik metode pengelompokan tersebut.

### 3. HASIL PENELITIAN DAN PEMBAHASAN

#### a. Deskripsi Data

sebesar 3702 jiwa/km<sup>2</sup>, dengan laju pertumbuhan penduduknya sebesar 1.24%. Selain itu persentase angkatan kerja sebesar 67.47%. Rata-rata pengeluaran perkapita sebesar Rp. 765951.7778. Angka harapan hidupnya rata-rata 71 tahun, sedangkan lama sekolah rata-rata 7.91 tahun.

#### b. Pengujian Asumsi Multikolinieritas

Sebelum dilakukan pengelompokan, dilakukan uji asumsi Multikolinieritas untuk mengetahui ukuran kemiripan apa yang digunakan.

Pengujian multikolinieritas didapat hasil berdasarkan persamaan (5) diperoleh sebagai berikut: nilai VIF dari variabel-variabel :  $KP = 4.557$  ;  $AK = 4.068$  ;  $LPP = 1.667$  ;  $PP = 2.919$  ;  $AHH = 3.734$  dan  $RLS = 3.707$  , Nampak terlihat bahwa nilai VIF untuk setiap variabel bernilai kurang dari 10. Sehingga dapat disimpulkan bahwa tidak terjadi multikolinieritas pada variabel-variabel tersebut. Oleh karena itu, dalam melakukan pengklasteran dapat menggunakan jarak *Euclidean*.

### c. Pembentukan Klaster *K-Means*

Hasil pengklasteran dengan menggunakan metode *K-means* adalah sebagai berikut:

#### 1) Jumlah Anggota Klaster

**Tabel 3.1 Jumlah anggota pada setiap klaster *k-means***

Klaster	Jumlah Anggota
1	8
2	19
Jumlah	27

Berdasarkan Tabel 3.2 hasil pengklasteran didapat 2 klaster dengan jumlah anggota pada klaster pertama 8 kab/kota, klaster kedua adalah 19 kab/kota di Jawa Barat.

#### 2) Karakteristik Klaster

Interpretasi karakteristik dari setiap klaster yang terbentuk adalah sebagai berikut:

**Tabel 3.2 Rata-rata variable pada klaster**

Variabel	Klaster-1	Klaster-2	Variabel	Klaster-1	Klaster-2
KP	8917.25	1412.725	PP	1075165	635756.9
AK	69.85	66.47211	AHH	72.89375	70.73947
LPP	2.011375	0.920105	RLS	9.88	7.081053

#### a) Klaster satu

Berdasarkan tabel di atas menunjukkan bahwa pada klaster satu kepadatan penduduk, angkatan kerja , laju pertumbuhan penduduk, pengeluaran riil perkapita, angka harapan hidup dan rata-rata lama sekolah memiliki nilai yang tinggi di banding klaster kedua. Sedangkan Anggota klaster satu antara lain Kab. Bekasi, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, dan Kota Cimahi ( Melalui Pembentukan cluster).

#### b) Klaster dua

Berdasarkan tabel di atas menunjukkan bahwa pada klaster dua memiliki nilai yang rendah dibanding klaster satu ,yang beranggotakan 19 Kab/kota. Untuk penentuan pengklusteran diperoleh melalui pembentukan Klaster *k-Means*.



#### d. Uji Variance Klaster K-Mean

Berdasarkan persamaan pada (10), (11), dan (12) didapat hasil sebagai berikut :

3.3 Tabel Nilai Varians

Variabel	$\sigma_b^2$	$\sigma_w^2$	$\sigma^2$
IPM	8.89129823	49.19833027	0.180723577
KP	12752250.81	32832789.98	0.388399853
AK	2.882911947	6.652021211	0.433388869
LPP	1.037981263	0.694265564	1.495078132
PP	48641924206	1.12563E+11	0.432129893
AHH	1.63268031	2.705603894	0.603443953
RLS	0.701249415	4.56720879	0.153540039

Berdasarkan tabel 3.3 terlihat bahwa nilai  $\sigma^2$  terbesar variabel LPP yaitu sebesar 1.495078132. Sehingga variable laju pertumbuhan penduduk adalah variabel yang memiliki perbedaan paling berarti pada klaster 1 dan klaster 2 serta merupakan variabel yang berpengaruh terhadap pembentukan klaster.

#### e. Analisis Diskriminan

Untuk mengetahui hasil ketepatan klasifikasi dengan *k-means* dilakukan analisis diskriminan dengan langkah-langkah berikut:

##### 1) Pengujian Asumsi

Sebelum melakukan analisis diskriminan terlebih dahulu melakukan pengujian asumsi-asumsi diantaranya:

##### a) Uji Normal Multivariate

Hasil pengujian asumsi normal didapat output dengan bantuan SPSS 19.0, ditunjukkan dengan plot Jarak mahalanobis dan Chi Kuadrat, diperoleh  $\rho_{hitung} = 0,992$  dengan nilai signifikansi sebesar 0.000, maka dengan demikian keduanya menunjukkan signifikansi kenormalan pada taraf 5%.

##### b) Uji Kehomogenan Matriks Kovarian Kedua Kelompok

Pengujian kehomogenan matriks kovarian gabungan kedua kelompok dengan menggunakan statistic uji Box'M ( $MC^{-1}$ ) berdasarkan persamaan (7) dan (8) diperoleh nilai  $MC^{-1} = 54.165$  dan dibandingkan dengan persamaan (9) dan (10) diperoleh nilai signifikansi sebesar 0.037, yang menunjukkan bahwa asumsi kehomogenan matriks kovarian dipenuhi pada taraf 5%. Dengan demikian analisis diskriminan, sesuai untuk mengelompokkan kabupaten/kota kedalam klaster satu dan klaster dua.

##### c) Uji Perbedaan Vektor Rata-Rata

Pada pengujian perbedaan vektor rata-rata dengan menggunakan SPSS 17.0, diperoleh hasil sebagai berikut:

4.4 Tabel Wilks' Lambda

Variabel	Wilks' Lambda	Sig.	Variabel	Wilks' Lambda	Sig.
KP	.410	.000	PP	.619	.001
AK	.514	.000	AHH	.265	.000
LPP	.733	.006	RLS	.439	.000

Pada tabel di atas menunjukkan bahwa semua variabel memiliki nilai signifikansi yang kurang dari 5%, ini berarti vektor rata-rata dari masing-masing variabel memiliki perbedaan secara nyata, sehingga asumsi ini terpenuhi.

## 2) Menghitung Ketepatan Klasifikasi

Tabel 3.5 Tabel Klasifikasi *k-means*

D	Prediksi		Total
	1	2	
Aktual	1	7	8
	2	0	19

Berdasarkan persamaan (3), maka didapat persentase ketepatan klasifikasi kluster *k-means* sebagai berikut :

$$CCR = \frac{7 + 19}{27} \times 100\% = 96.3\%$$

Berdasarkan persamaan (4), maka didapat persentase ketepatan klasifikasi kluster *k-means* sebagai berikut :

$$APER = \frac{0 + 1}{27} \times 100\% = 3.7\%$$

Sehingga perhitungan tersebut dapat terlihat bahwa nilai ketepatan klasifikasi *k-means* tinggi dengan tingkat kesalahan 3.7%.

## 4. SIMPULAN

Pada metode kluster *k-means* nilai CCR sebesar 96.3% ini dan tingkat kesalahan sebesar 3,7 % menunjukkan bahwa klasifikasi yang terbentuk, mempunyai tingkat ketepatan yang sangat tinggi. Pengelompokan 27 Kabupaten/Kota di Provinsi Jawa Barat berdasarkan indikator kesejahteraan rakyat tahun 2014 dapat dibentuk 2 kluster. Kluster pertama terdiri dari delapan Kab/kota yakni : Kab. Bekasi, Kota Bogor, Kota Cimahi, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi dan Kota Depok, sedangkan sisanya kluster kedua terdiri dari Sembilan belas Kab/kota.

Secara umum kesejahteraan masyarakat Provinsi Jawa Barat berdasarkan Indikator Kesejahteraan Rakyat Tahun 2014 masih belum merata. Hal ini dapat dilihat pada kluster pertama dari kepadatan penduduk, angkatan kerja, laju pertumbuhan penduduk, rata-rata pengeluaran perkapita, angka harapan hidup, dan rata-rata lama sekolah menunjukkan keunggulan dibanding kluster dua. Sehingga, hasil pengklasteran dengan metode *k-means* dapat dikatakan bahwa kabupaten/kota yang ada pada kluster kedua lebih membutuhkan bantuan pemerintah di banding dengan kabupaten/kota pada kluster yang pertama.

**5. DAFTAR PUSTAKA**

- Basri , Faizal dan Haris Munandar( 2009) .Lanskap Ekonomi Indonesia, Jakarta:
- Bidang Statistik Sosial, BPS Provinsi Jabar. (2015). *Indikator Kesejahteraan*
- Dillon,William and Matthew.(1984). *Multivariate Analysis Methods and Applications* .John Wiley & Sons .New York.
- Dumairy.(1996) *Perekonomian Indonesia* .Penerbit Erlangga . Jakarta
- Febriyana. (2011). *Analisis Klaster K-means dan K-Median pada Data Indikator Kemiskinan (Studi Kasus Data Indikator Kemiskinan Kabupaten Indonesia Tahun 2009)*. Universitas Islam Negeri Syarif Hidayatullah, Jakarta.
- Johanes ,Supranto.( 2004) . *Analisis Multivariate ,arti dan Interpretasi* . Rineka Cipta Jakarta.
- Johnson, Richard A. dan Dean W Wichern. (1998). *Applied Multivariate Statistical Analysis*.Edisi keempat. New York: Prentice-Hall International, Inc.
- Rustam, Supardjo.(1993).*Pembangunan Nasional Untuk Kesejahteraan Rakyat*. Perpustakaan Nasional RI, Jakarta.