# PERBANDINGAN INDEKS VALIDITAS GAP STATISTIC DAN INDEKS VALIDITAS SILHOUETTE PADA K-MEANS CLUSTER

(Studi Kasus pada Data Penyusunan Kinerja Makro Ekonomi dan Sosial tentang Pembangunan Manusia Jawa Timur)

BRAWIUNA Oleh: DEWI RATNASARI 0710950008-95



PROGRAM STUDI STATISTIKA JURUSAN MATEMATIKA FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM UNIVERSITAS BRAWIJAYA **MALANG** 2011



# PERBANDINGAN INDEKS VALIDITAS GAP STATISTIC DAN INDEKS VALIDITAS SILHOUETTE PADA K-MEANS CLUSTER

(Studi Kasus pada Data Penyusunan Kinerja Makro Ekonomi dan Sosial tentang Pembangunan Manusia Jawa Timur)

#### **SKRIPSI**

Sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains dalam Bidang Statistika

> Oleh: DEWI RATNASARI 0710950008-95



PROGRAM STUDI STATISTIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2011

# LEMBAR PENGESAHAN SKRIPSI PERBANDINGAN INDEKS VALIDITAS GAP STATISTIC DAN INDEKS VALIDITAS SILHOUETTE

PADA K-MEANS CLUSTER

(Studi Kasus pada Data Penyusunan Kinerja Makro Ekonomi dan Sosial tentang Pembangunan Manusia Jawa Timur)

# Oleh: DEWI RATNASARI NIM. 0710950008-95

BRAWING A Setelah dipertahankan di depan Majelis Penguji pada tanggal Oktober 2011 dan dinyatakan memenuhi syarat untuk memperoleh gelar Sarjana Sains dalam bidang Statistika

Pembimbing I,

Pembimbing II,

Prof. Dr. Ir. Waego Hadi N. NIP. 19521207 197903 1 003

Ir. Heni Kusdarwati, MS. NIP. 19611208 198701 2 001

Mengetahui, Ketua Jurusan Matematika Fakultas MIPA Universitas Brawijava

Dr. Abdul Rouf Al-Ghofari, M.Sc NIP. 19670907 199203 1 001

#### LEMBAR PERNYATAAN

## Saya yang bertanda tangan di bawah ini:

Nama : DEWI RATNASARI

NIM : 0710950008 - 95 Program Studi : STATISTIKA

Penulisan Skripsi berjudul :

Perbandingan Indeks Validitas Gap Statistic dan Indeks Validitas Silhouette pada K-Means Cluster

(Studi Kasus pada Data Penyusunan Kinerja Makro Ekonomi dan Sosial tentang Pembangunan Manusia Jawa Timur)

#### Dengan ini menyatakan bahwa:

- 1. Isi dari Skripsi yang saya buat adalah benar-benar karya sendiri dan tidak menjiplak karya orang lain selain namanama yang termaktub di isi dan tertulis di daftar pustaka dalam Skripsi.
- 2. Apabila di kemudian hari ternyata Skripsi yang saya tulis terbukti hasil jiplakan, maka saya akan bersedia menanggung segala risiko yang akan saya terima.

Demikian pernyataan ini dibuat dengan segala kesadaran.

Malang, September 2011 Yang menyatakan,

(DEWI RATNASARI) NIM. 0710950008

# PERBANDINGAN INDEKS VALIDITAS GAP STATISTIC DAN INDEKS VALIDITAS SILHOUETTE PADA K-MEANS CLUSTER

(Studi Kasus pada Data Penyusunan Kinerja Makro Ekonomi dan Sosial tentang Pembangunan Manusia Jawa Timur)

#### **ABSTRAK**

K-Means Cluster merupakan metode pengelompokan data non hirarki yang berusaha mempartisi data ke dalam bentuk satu atau lebih cluster, dengan mengasumsikan banyaknya kelompok diketahui dan telah didefinisikan. Validitas *cluster* perlu dilakukan karena belum ada dasar yang kuat mengenai jumlah kelompok terbaik dalam penentuan iumlah kelompok pada analisis *cluster*. Tibshirani *et al.* (2001) dan Pramono (2005) telah melakukan penelitian tentang berbagai metode penentuan jumlah kelompok optimal dengan menggunakan indeks validitas. Pada penelitian ini digunakan indeks validitas Gap Statistic dan indeks validitas Silhouette dalam menentukan jumlah optimum kelompok. Tujuan dari penelitian ini adalah membandingkan indeks validitas Gap Statistic dan indeks validitas Silhouette berdasarkan nilai Cluster Tightness Measure (CTM). Data yang digunakan dalam penelitian ini adalah data Penyusunan Kinerja Makro Ekonomi dan Sosial Jawa Timur tentang Pembangunan Manusia dari Tahun 2005-2008 dengan 4 variabel penelitian pada sampel berukuran 38. Hasil penelitian yang diperoleh dari keseluruhan hasil pengelompokan menunjukkan bahwa Indeks validitas Silhouette menghasilkan nilai CTM terkecil. Sehingga dapat disimpulkan bahwa indeks validitas Silhouette merupakan indeks validitas yang lebih baik dalam penentuan jumlah cluster optimum dibandingkan indeks validitas Gap Statistic pada metode K-means Cluster. Perbandingan hasil pengelompokan pada indeks validitas Silhouette dengan k=3 dari 2005 ke Tahun 2008 menunjukkan pergeseran hasil pengelompokan. Pemerintah harus mengarahkan perhatian pada wilayah yang masuk pada kelompok 3. Karena, wilayah yang masuk pada kelompok 3 adalah wilayah yang memiliki indeks kesehatan, indeks pendidikan, daya beli dan Indeks Pembangunan Manusia (IPM) yang paling rendah dibandingkan kelompok yang lain.

Kata kunci : K-Means Cluster, Gap Statistic, Silhouette, Cluster Tightness Measure (CTM).

# COMPARISON OF GAP STATISTIC VALIDITY INDEX AND SILHOUETTE VALIDITY INDEX ON K-MEANS CLUSTER

(Case Study of Macro Economic and Social Performance on Human Development Data in East Java)

#### **ABSTRACT**

K-Means Cluster is a non-hierarchical method of grouping data that seek to divide the data into the form of one or more clusters, assuming that the number of group are known and has been defined. There is no solid basis on the number of the best group, therefore, the validity of the cluster needs to be done. Tibshirani et al. (2001) dan Pramono (2005) analyzed a variety of methods of determining the optimal number of groups using the validity index. This research will use the Gap Statistic validity index and Silhouette validity index in determining the optimum number of groups. The purpose of this study is to compare the Gap Statistic validity index and Silhouette validity index based on the Cluster Tightness Measure (CTM). This study uses Macro Economic and Social Performance on Human Development data in East Java of the Year 2005-2008 which consists of 4 variables with 38 samples. The results show that the clustering based on Silhouette validity index yields the smallest value of the CTM. Therefore Silhouette validity index is better in determining the optimum number of cluster than Gap Statistic validity index on Kmeans cluster method. Comparison of clustering results of the Silhouette validity index with k=3 of the Year 2005 to Year 2008 showed a shift in the grouping. Governments must give more attention to the areas that are inside group 3. Because, the regions in group 3 are the regions with lowest health index, education index, purchasing power and Human Development Index (HDI) than other groups.

Keywords: K-Means Cluster, Gap Statistic, Silhouette, Cluster Tightness Measure (CTM).

#### KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat, karunia, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Perbandingan Indeks Validitas *Gap Statistic* dan Indeks Validitas *Silhouette* pada *K-Means Cluster* (Studi Kasus pada Data Penyusunan Kinerja Makro Ekonomi dan Sosial tentang Pembangunan Manusia Jawa Timur)". Dalam pelaksanaan dan penyusunan skripsi ini, banyak pihak yang telah membantu baik berupa bimbingan, dukungan, dan saran. Oleh karena itu, penulis mengucapkan terimakasih kepada:

- 1. Bapak Prof. Dr. Ir. Waego Hadi N selaku dosen pembimbing I dan Ibu Ir. Heni Kusdarwati, M.S. selaku dosen pembimbing II yang telah memberikan bimbingan, dukungan, serta saran.
- 2. Ibu Dr. Rahma Fitriani S.Si., M.Sc selaku dosen penguji yang telah memberikan saran dan kritik.
- 3. Bapak Dr. Abdul Rouf A., M.Sc selaku Ketua Jurusan Matematika Fakultas MIPA Universitas Brawijaya.
- 4. Ibu, Bapak, Kakak, dan seluruh anggota keluarga atas doa, kasih sayang, serta dukungan moril dan materil yang tidak pernah putus.
- 5. Teman-teman Statistika 2007 atas dukungan, semangat, dan persahabatan yang tidak akan terlupakan.
- 6. Teman-teman kos Mertojoyo Selatan A6 atas motivasi, perhatian, dan kebersamaan yang tidak akan terlupakan
- 7. Karyawan tata usaha dan semua pihak yang telah membantu dalam proses penyelesaian skripsi ini

Penulis menyadari bahwa skripsi ini masih terdapat kekurangan. Oleh karena itu, penulis menerima saran dan kritik yang membangun demi kesempurnaan skripsi ini. Semoga skripsi ini dapat bermanfaat bagi berbagai pihak.

Malang, September 2011

Penulis

# DAFTAR ISI

Ha	laman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN	iii
ABSTRAK	iv
ABSTRACTKATA PENGANTAR	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	X
DAFTAR LAMPIRAN	xi
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan penelitian	3
1.4 Batasan Masalah	3
1.5 Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA	
2.1 Korelası	5
2.2 Analisis Komponen Utama	6
2.3 Analisis Cluster	9
2.3.1 Metode <i>K-Means</i>	11
2.4 Indeks Validitas	13
2.4.1 Gap Statistic	13
2.4.2 Silhouette	15
2.5 Ukuran Kedekatan	17
2.6 Cluster Tightness Measure (CTM)	18
2.7 Sumber Daya Manusia	19
2.7.1 Indeks Pembangunan Manusia	19
2.7.2 Angka Harapan Hidup	20
2.7.3 Tingkat Pendidikan	20
2.7.4 Standar Layak Hidup	20

	lamar
BAB III METODE PENELITIAN	
3.1 Data Penelitian	23
3.2 Metode Penelitian	23
3.2.1 Tahap persiapan	23
3.2.2 Tahap Analisis Data	24
3.3 Diagram Alir	25
BAB IV HASIL PEMBAHASAN	
4.1 Uji Koefisien Korelasi	29
4.2 Analisis Komponen Utama	32
4.2.1 Hasil Analisis Komponen Utama Data 1	32
4.2.2 Hasil Analisis Komponen Utama Data 2	33
4.2.3 Hasil Analisis Komponen Utama Data 3	34
4.2.4 Hasil Analisis Komponen Utama Data 4	34
4.3 Analisis <i>K-Means Cluster</i> dengan Indeks Validitas <i>Gap</i>	
Statistic	35
4.4 Analisis <i>K-Means Cluster</i> dengan Indeks Validitas	
Silhouette	37
4.5 Perbandingan Indeks Validitas <i>Gap Statistic</i> dan Indeks	
Validitas Silhouette	39
4.6 Perbandingan Hasil Pengelompokan dari tahun ke tahun	41
4.0 Terbandingan Trash Tengerompokan dari tahun ke tahun	71
BAB V KESIMPULAN DAN SARAN	
5.1 Kesimpulan	49
5.2 Saran	50
	50
DAFTAR PUSTAKA	51
LAMPIRAN	55

# DAFTAR TABEL

	Ha	alaman
Tabel 4.1	Hasil Pengujian Koefisien Korelasi Data 1	29
Tabel 4.2	Hasil Pengujian Koefisien Korelasi Data 2	30
Tabel 4.3	Hasil Pengujian Koefisien Korelasi Data 3	30
Tabel 4.4	Hasil Pengujian Koefisien Korelasi Data 4	31
Tabel 4.5	Nilai Indeks Validitas Gap statistic (gap)	36
Tabel 4.6	Nilai Indeks Validitas Silhouette (S)	37
Tabel 4.7	Nilai CTM Indeks Validitas Gap statistic dan	
	Silhouette	39
Tabel 4.8	Hasil Pengelompokan dengan Indeks Validitas	
	<i>Silhouette</i> untuk <i>k</i> =3	42

# DAFTAR GAMBAR

		tataman
Gambar 3.1	Diagram Alir metode Penelitian	26
Gambar 3.2	Skema Algoritma K-Means Cluster	27



# DAFTAR LAMPIRAN

	H	Halaman
Lampiran 1.	Data Penelitian	. 55
Lampiran 2.	Uji Koefisien Korelasi	. 63
Lampiran 3.	Hasil Analisis Komponen Utama	. 65
Lampiran 4.	Nilai indeks validitas <i>Gap Statistic, diffu</i> da <i>cluster membership</i> hasil output <i>software</i>	ın
	R.2.13.0	. 73
Lampiran 5.	Nilai indeks validitas <i>Silhouette</i> dan <i>cluster membership</i> hasil output <i>software</i>	4
	R.2.13.0	. 77
Lampiran 6.	Hasil Pengelompokan Tahun 2005,	
	Tahun 2006, Tahun 2007 dan Tahun 2008 metode <i>K-means</i> pada Indeks Validitas	
	Silhouette dengan k=3	. 81



#### BAB I PENDAHULUAN

### 1.1 Latar Belakang

Analisis *cluster* merupakan suatu teknik multivariat yang digunakan untuk mengelompokkan objek-objek ke dalam beberapa *cluster* berdasarkan kesamaan karakteristik di antara objek-objek tersebut. Objek yang berada dalam kelompok yang sama relatif lebih homogen daripada objek yang berada pada kelompok yang berbeda. Setiap *cluster* memiliki karakteristik yang berbeda dengan *cluster* yang lain.

*K-means* merupakan metode *cluster* yang bertujuan untuk mengelompokkan data sedemikian hingga jarak tiap-tiap data ke pusat kelompok dalam satu *cluster* minimum. Metode *K-means* memiliki beberapa kelebihan dibandingkan metode hirarki, karena hasilnya tidak mudah terpengaruh *outlier*, jarak pengukuran yang digunakan dan penyertaan variabel yang tidak sesuai (Hair, 1998).

Salah satu masalah utama dalam analisis *cluster* adalah penentuan jumlah optimal kelompok dalam *cluster*. Setelah melakukan analisis *cluster*, perlu dilakukan validitas *cluster* karena belum ada dasar yang kuat mengenai jumlah kelompok terbaik pada penentuan jumlah kelompok dalam analisis *cluster*. Validitas cluster merupakan prosedur untuk mengevaluasi hasil analisis *cluster* secara kuantitatif sehingga dihasilkan kelompok optimum. Kelompok optimum yaitu kelompok yang mempunyai jarak yang padat antar individu dalam kelompok dan terisolasi dari kelompok lain dengan baik (Dubes dan Jain, 1988).

Gap statistik merupakan suatu teknik penentuan jumlah optimal kelompok dari suatu himpunan data. Teknik ini dapat digunakan pada hasil pengelompokan dari berbagai metode pengelompokan, misal: *K-means* dan metode hirarki (Tibshirani *et al*, 2001).

Silhouette dapat digunakan untuk mengevaluasi kevalidan hasil dari suatu analisis cluster serta dapat digunakan untuk menentukan jumlah cluster yang sesuai. Silhouette tidak dipengaruhi oleh metode pengelompokan yang digunakan. Jadi, silhouette dapat digunakan untuk meningkatkan hasil dari analisis cluster atau untuk membandingkan hasil dari metode cluster yang berbeda yang diaplikasikan pada data yang sama (Rousseeuw, 1987).

Tibshirani *et al.* (2001) dalam penelitiannya membandingkan beberapa indeks validitas antara lain Calinski-harabaz, Hartigan, Krzanowski-Lai dan Silhouette, yang selanjutnya dibandingkan dengan indeks validitas *Gap Statistic*. Penelitian ini dilakukan dengan membangkitkan sejumlah himpunan data dengan beberapa tetapan jumlah kelompok yang telah ditentukan. Hasil penelitian ini, indeks validitas *Gap Statistic* secara umum paling baik dalam menentukan kelompok optimum. Tetapi, estimasi indeks validitas *Gap statistic* akan baik pada data yang terkelompok dengan baik berdasarkan kriteria tertentu sehingga objek-objek dalam kelompok mempunyai ragam yang lebih kecil dibandingkan ragam antar kelompok.

Penelitian yang dilakukan Pramono (2005), adalah membandingkan prosedur pengelompokan Hartigan, Krzanowski-Lai, Silhouette dan gap statistic. Hasil dari penelitian ini, dengan menggunakan metode K-means cluster indeks validitas Silhouette dan Gap Statistic sama baiknya dalam mengestimasi jumlah kelompok optimum. Tetapi jika jarak antar kelompok lebih lebar maka indeks validitas Silhouette memberikan hasil estimasi yang lebih baik daripada indeks validitas Gap Statistic. Hasil estimasi dengan indeks validitas silhouette memiliki akurasi 100% dan gap statistic memiliki akurasi sebesar 90%.

Walaupun Tibshirani et al. (2001) menyimpulkan bahwa indeks validitas Gap Statistic memberikan hasil pengelompokan optimum yang terbaik akan tetapi Pramono (2005) memberi indikasi bahwa indeks validitas silhouette memberikan hasil estimasi yang lebih baik dalam menentukan jumlah optimum kelompok. Oleh sebab itu, terlepas dari tipe data yang digunakan penelitian ini bertujuan untuk membandingkan indeks validitas Gap Statistic dan indeks validitas Silhouette. Kedua indeks validitas tersebut, dibandingkan nilai prosentase ketepatan estimasi masing-masing terhadap metode K-means Cluster dengan diterapkan pada data Penyusunan Kinerja Makro Ekonomi dan Sosial Jawa Timur tentang Pembangunan Manusia dari Tahun 2005 sampai Tahun 2008.

Sumber Daya Manusia (SDM) yang berkualitas merupakan aset yang penting bagi pembangunan di berbagai aspek kehidupan masyarakat. SDM yang berkualitas dapat tercipta apabila setiap warga masyarakat dapat memperoleh pelayanan pendidikan dan kesehatan yang memadai, merata dan bermutu. Indeks Pembangunan Manusia (IPM) merupakan ukuran capaian pembangunan manusia berbasis

sejumlah komponen dasar kualitas hidup. Terdapat tiga komponen yang memiliki keterkaitan yang akan digunakan untuk melihat kemajuan manusia yaitu berdasarkan dimensi kesehatan yang diukur dari rata-rata usia harapan hidup, dimensi pendidikan yang diukur dari rata-rata lama sekolah dan angka melek huruf, serta dimensi ekonomi yang diukur dari tingkat kehidupan yang layak (kesejahteraan) secara keseluruhan.

#### 1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah :

- 1. Bagaimana hasil estimasi jumlah kelompok optimum menggunakan indeks validitas *Gap Statistic* pada metode *K-means*?
- 2. Bagaimana hasil estimasi jumlah kelompok optimum menggunakan indeks validitas *Silhouette* pada metode *K-means*?
- 3. Indeks validitas apa yang terbaik untuk menentukan kelompok optimum?
- 4. Bagaimana perubahan kelompok pada data Penyusunan Kinerja Makro ekonomi dan Sosial Jawa Timur tentang Pembangunan Manusia dari Tahun 2005 ke Tahun 2008?

# 1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

- 1. Melakukan pendugaan jumlah kelompok optimum menggunakan indeks validitas *Gap Statistic* pada metode *K-means*.
- Melakukan pendugaan jumlah kelompok optimum menggunakan indeks validitas Silhouette pada metode Kmeans.
- 3. Membandingkan hasil estimasi indeks validitas *Gap Statistic* dan *Silhouette* pada nilai *k* yang sama untuk mengetahui indeks validitas terbaik.
- 4. Melakukan analisis terhadap perubahan kelompok data Penyusunan Kinerja Makro ekonomi dan Sosial Jawa Timur tentang Pembangunan Manusia dari Tahun 2005 ke Tahun 2008.

#### 1.4 Batasan Masalah.

Batasan masalah pada penelitian ini adalah:

- 1. Metode pengelompokan yang digunakan adalah metode *K-means cluster*.
- 2. Jarak yang digunakan adalah jarak Euclid.
- 3. Indeks validitas yang digunakan adalah indeks validitas *Gap Statistic* dan *Silhouette*.
- 4. Penentuan indeks validitas terbaik berdasarkan pada nilai *Cluster Tightness Measure* (CTM).

#### 1.5 Manfaat Penelitian

Manfaat dari hasil penelitian ini adalah:

- 1. Dapat membantu pengguna analisis kelompok dalam menentukan indeks validitas terbaik yang akan digunakan dalam memprediksi kelompok optimum.
- Dapat mengetahui pergeseran keanggotaan kelompok dari Tahun 2005 ke Tahun 2008 dari data Penyusunan Kinerja Makro Ekonomi dan Sosial Jawa Timur tentang Pembangunan Manusia dari Tahun 2005 sampai Tahun 2008.
- 3. Dapat mengetahui wilayah mana yang masuk pada kategori memiliki pembangunan manusia rendah, sedang maupun tinggi. Sehingga dapat memberikan saran kepada Pemerintah Provinsi Jawa Timur bahwa harus memfokuskan perhatian pada wilayah yang masuk pada wilayah dengan kategori pembangunan manusia yang masih rendah.

### BAB II TINJAUAN PUSTAKA

#### 2.1 Korelasi

Korelasi adalah metode untuk mengetahui tingkat keeratan hubungan dua peubah atau lebih yang digambarkan oleh besarnya koefisien korelasi. Koefisien korelasi adalah koefisien yang menggambarkan tingkat keeratan hubungan antar dua peubah atau lebih. Besaran dari koefisien korelasi tidak menunjukan hubungan sebab akibat antara dua variabel, tetapi hanya menggambarkan keterkaitan linier antar variabel (Walpole, 1992).

Ukuran hubungan linier antara dua variabel *X* dan *Y* dirumuskan sebagai berikut:

$$r_{x,y} = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right) \left(\sum_{i=1}^{n} y_i\right)}{\sqrt{\left(n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right) \left(n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y\right)^2\right)}}$$
(2.1)

di mana:

 $r_{x,y}$ : koefisien korelasi variabel X dan Y

n: banyaknya data

Uji koefisien korelasi dilakukan untuk mengetahui adanya korelasi antar dua peubah dengan hipotesis sebagai berikut:

 $H_0$ : Tidak ada korelasi

 $H_1$ : Ada korelasi

Dengan statistik uji sebagai berikut:

$$t_{hit} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$
 (2.2)

di mana:

r : koefisien korelasin : banyaknya data

Pengujian korelasi dilakukan menggunakan sebaran t dengan derajat bebas (n-2) dan taraf nyata  $(\alpha)$  tertentu dalam menentukan daerah kritik yang menjadi dasar pengambilan keputusan. Ketentuan dalam uji koefisien korelasi ini adalah apabila  $|t_{hit}| \leq t_{tabel}$  maka terima  $H_0$ , artinya kedua variabel saling bebas, jika $|t_{hit}| > t_{tabel}$  maka tolak  $H_0$ , artinya ada hubungan antara dua variabel (Walpole, 1992).

Selain menggunakan statistik uji, penarikan kesimpulan suatu hipotesis juga bisa dilakukan dengan menggunakan peluang signifikan (*p-value*). *P-value* didefinisikan sebagai peluang untuk memperoleh hasil paling sedikit sama ekstrim dengan titik data yang diberikan. Misalnya titik data yang diberikan adalah statistik uji *t* yang didapatkan dari persamaan (2.2). Jika statistik uji *t* tersebut mengikuti *studentized distribution* atau distribusi T maka *p-value* didapatkan dari:

$$p = P(T \ge t_{hit}) \tag{2.3}$$

di mana p adalah p-value, nilai kritis pada distribusi T atau studentized distribution. Untuk tingkat kepercayaan  $\alpha$ , perbandingan p dengan  $\alpha$  memungkinkan untuk membuat suatu kesimpulan bagi ditolak atau tidaknya suatu hipotesis nol yaitu jika  $p \le \alpha$  maka  $H_0$  ditolak dan  $p > \alpha$  maka  $H_0$  diterima. (Dodge, 2008)

Nilai koefisien korelasi  $(r_{x,y})$  berkisar antara -1 sampai 1. Hubungan linier sempurna terjadi jika nilai  $r_{x,y} = 1$  atau  $r_{x,y} = -1$ . Bila  $r_{x,y}$  mendekati 1 atau -1 hubungan antara kedua peubah kuat dan dikatakan terdapat korelasi yang tinggi antar keduanya. Akan tetapi bila  $r_{x,y}$  mendekati nol, hubungan linier antar X dan Y sangat lemah. Untuk  $r_{x,y} = 0$ , diartikan tidak terdapat hubungan linier antara antar X dan Y (Walpole, 1992).

### 2.2 Analisis Komponen Utama

Analisis Komponen Utama (AKU) merupakan suatu metode statistika yang digunakan untuk mentransformasi sekumpulan variabel menjadi variabel baru di mana variabel yang dihasilkan tidak saling berkorelasi dan merupakan kombinasi linier variabel asal. Variabel yang dihasilkan dinamakan komponen utama (Afifi dan Clark, 1990)

Menurut Gasperz dalam Kirana (2008) metode AKU pertama kali diperkenalkan oleh Harold Hotteling (1933), dan pada dasarnya

bertujuan bertujuan untuk menerangkan struktur ragam peragam melalui kombinasi linier dari peubah-peubah penjelas. Meskipun dari p peubah asal dapat diturunkan menjadi p peubah baru untuk menerangkan ragam total sistem, namun seringkali ragam total itu dapat diterangkan secara baik oleh sejumlah kecil (q) peubah baru.

Metode AKU mentransformasi vektor kolom peubah asal X menjadi vektor kolom peubah baru Y di mana  $q \leq p$ . AKU didefinisikan sebagai berikut :

$$Y_i = \sum_{j=1}^p a_{ij} X_j \tag{2.4}$$

di mana  $1 \le I \le q, \ 1 \le j \le p$  dan  $Y_1, Y_2, ..., Y_q$  saling bebas. Vektor peubah baru  $Y_q$  menjelaskan sebesar mungkin proporsi keragaman vector peubah asal.

AKU mentransformasi vektor peubah asal  $\mathbf{X} = (X_1, X_2, ..., X_p)$  menjadi vektor peubah baru  $\mathbf{Y} = (Y_1, Y_2, ..., Y_p)$ . Persamaan pada Analisis Komponen Utama adalah sebagai berikut:

$$Y_{1} = \mathbf{a}_{1}^{'} \mathbf{X} = a_{11} X_{1} + a_{12} X_{2} + \dots + a_{1p} X_{p}$$

$$Y_{2} = \mathbf{a}_{2}^{'} \mathbf{X} = a_{21} X_{1} + a_{22} X_{2} + \dots + a_{2p} X_{p}$$

$$\vdots$$

$$Y_{p} = \mathbf{a}_{p}^{'} \mathbf{X} = a_{p1} X_{1} + a_{p2} X_{2} + \dots + a_{pp} X_{p}$$

$$(2.5)$$

di mana  $a_1, a_2, ..., a_p$  adalah vektor koefisien untuk setiap komponen utama yang bersesuaian. Rumus ragam dan peragam untuk peubah Y adalah:

$$Var(Y_i) = a_i' \Sigma a_i$$
  $i = 1, 2, ... p$  (2.6)

$$Cov(Y_i) = a_i' \Sigma a_k$$
  $i, k = 1, 2, ..., p$  (2.7)

di mana  $\Sigma$  adalah matriks ragam peragam peubah asal.

Komponen utama merupakan kombinasi linier dengan ragam maksimum,  $\mathbf{a}_i' \mathbf{\Sigma} \mathbf{a}_i$  adalah ragam komponen utama yang akan mencapai nilai maksimum dengan syarat  $\mathbf{a}_i' \mathbf{a}_i = 1$ .

Permasalahan ini dapat diselesaikan menggunakan fungsi Lagrange dengan maksimumkan ragam  $a_i \Sigma a_i$ , dengan batasan:

 $\mathbf{a}_i'\mathbf{a}_i = 1$  atau  $\mathbf{a}_i'\mathbf{a}_i - 1 = 0$ . Sehingga didapatkan fungsi Lagrange sebagai berikut:

$$L = \mathbf{a}_{i}^{'} \mathbf{\Sigma} \mathbf{a}_{i} - \lambda_{i} (\mathbf{a}_{i}^{'} \mathbf{a}_{i} - 1)$$
 (2.8)

Kemudian L diturunkan terhadap  $a_i$  kemudian disamakan dengan nol maka diperoleh:

$$\frac{\partial L}{\partial \boldsymbol{a}} = 2\boldsymbol{a}_{i}\boldsymbol{\Sigma} - 2\lambda_{i}\boldsymbol{a}_{i} = 0$$

$$2(\boldsymbol{\Sigma} - \lambda_{i}\boldsymbol{I})\boldsymbol{a}_{i} = 0$$

$$(\boldsymbol{\Sigma} - \lambda_{i}\boldsymbol{I})\boldsymbol{a}_{i} = 0$$
(2.9)

Berdasarkan persamaan (2.9) dihasilkan jawaban nontrivial yang berarti hanya ada satu jawaban yang bersifat unik apabila matriks ( $\Sigma - \lambda_i I$ ) merupakan matriks singular (memiliki determinan sama dengan nol)

$$|(\mathbf{\Sigma} - \lambda_i \mathbf{I})| = 0 \tag{2.10}$$

Persamaan (2.10) akan menghasilkan nilai-nilai *Eigen* yaitu  $\lambda_1, \lambda_2, ..., \lambda_p$  dimana  $\lambda_1 > \lambda_2 > ... > \lambda_p > 0$ . Setiap nilai *Eigen*  $\lambda_i$  akan menentukan vektor *Eigen*  $\boldsymbol{a}_i$ .

Penentuan nilai *Eigen* yang akan digunakan pada komponen utama pertama yaitu:

$$(\mathbf{\Sigma} - \lambda_i \mathbf{I}) \mathbf{a}_i = 0$$

$$\mathbf{\Sigma} \mathbf{a}_i - \lambda_i \mathbf{I} \mathbf{a}_i = 0$$

$$\mathbf{\Sigma} \mathbf{a}_i = \lambda_i \mathbf{I} \mathbf{a}_i$$
(2.11)

Jika kedua sisi persamaan (2.11) dikalikan dengan  $\mathbf{a}_i'$  maka diperoleh persamaan :

$$\mathbf{a}_{i}^{'}\mathbf{\Sigma}\mathbf{a}_{i} = \lambda_{i} \tag{2.12}$$

Ragam setiap komponen utama bersesuaian dengan setiap nilai Eigen. Persamaan (2.11) menunjukan bahwa ragam komponen utama maksimum adalah nilai Eigen terbesar dari matriks  $\Sigma$ .

Proporsi keragaman total yang dijelaskan oleh komponen utama ke-*k* adalah:

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \times 100\% \tag{2.13}$$

Pada analisis komponen utama walaupun terbentuk p komponen utama yang menjelaskan keragaman total, tetapi biasanya sebagaian besar keragaman total bisa dijelaskan oleh k komponen utama di mana  $k \le p$  (Johnson dan Wichern, 2002).

Menurut Dillon dan Goldstein (1984), pendekatan yang dapat digunakan dalam penentuan komponen utama yang digunakan yaitu pendekatan akar karakteristik lebih besar dari satu. Pendekatan ini mengambil komponen utama yang mempunyai nilai eigen lebih besar dari satu. Selain itu pendekatan lain yang diberikan adalah pendekatan Scree Test. Pada pendekatan ini, akar karakteristik dari masingmasing komponen utama diplot berdasarkan urutan besarnya akar karakteristik. Dasar pemikirannya adalah bahwa hasil komponen utama sudah diurutkan berdasarkan besarnya akar karakteristik, sehingga komponen utama pertama akan muncul pertama, dan kemudian diikuti oleh komponen utama lainnya yang mempunyai proporsi keragaman yang lebih kecil. Menurut Johnson dan Wichern (2002), proporsi keragaman total dari k komponen utama yang bernilai lebih dari 80 atau 90 % menunjukkan bahwa k komponen utama tersebut dapat menggantikan variabel-variabel asli tanpa kehilangan banyak informasi.

Analisis komponen utama bisa digunakan sebagai analisis awal untuk analisis yang lain. Hasil analisis komponen utama bisa digunakan sebagai masukan pada analisis regresi berganda dan analisis kelompok (Johnson dan Wichern, 2002).

#### 2.3 Analisis Cluster

Analisis *cluster* adalah suatu metode dalam analisis peubah ganda yang bertujuan untuk mengelompokkan n satuan pengamatan ke dalam k kelompok dengan (k < n) berdasarkan p peubah, sehingga unit-unit pengamatan dalam suatu kelompok memiliki ciri-ciri yang lebih homogen dibandingkan unit pengamatan dalam kelompok lain.

Analisis *cluster* mengidentifikasi sekelompok objek berdasarkan kemiripan karakteristik tertentu yang dapat dipisahkan dengan kelompok lainnya. Pendefinisian kesamaan (homogenitas) kelompok yang ada sangat bergantung pada tujuan studi atau penelitian (Sharma, 1996).

Analisis *cluster* merupakan salah satu teknik analisis dalam data mining dimana *clustering* melakukan pengelompokan data berdasarkan kesamaan karakteristik data. Dengan kesamaan karakteristik pada sebuah kelompok, dapat diambil suatu informasi yang mempunyai arti dan berguna (Handriyadi *et al*, 2009)

Analisis *cluster* membantu penyederhanaan permasalahan dengan melakukan pengelompokan berdasarkan karakteristik peubah ke dalam sejumlah kelompok yang relatif lebih homogen. Kelompok merupakan kumpulan objek yang homogen atau memiliki kemiripan satu sama lain. Pengelompokan data dapat meningkatkan informasi dari data untuk mengurangi pengaruh proporsi dari ragam (Marriot, 1971).

Konsep dasar pengukuran analisis kelompok adalah konsep pengukuran jarak (distance) dan kesamaan (similarity). Distance adalah ukuran tentang jarak pisah antar objek sedangkan similarity adalah ukuran kedekatan. Konsep ini penting karena pengelompokan pada Analisis Cluster didasarkan pada kedekatan. Pengukuran jarak digunakan untuk data yang bersifat metrics, sedangkan pengukuran kesesuaian (matching type measure) digunakan untuk data yang bersifat kualitatif (Laboratorium Data Mining Jurusan Teknik Industri Fakultas Teknik Industri Fakultas Teknik Industri Universitas Islam Indonesia, 2009)

Analisis *cluster* terbagi menjadi dua metode utama, yaitu metode hirarki dan metode non hirarki atau biasa dikenal dengan metode partisi (penyekatan). Metode hirarki secara umum dibagi menjadi dua yaitu *agglomerative* (penggabungan) dan *divisive* (pembagian). Sementara metode non hirarki algoritmanya disusun dengan cara membagi data menjadi k kelompok, dimana besarnya nilai k ditentukan sendiri oleh penggunanya, metode ini biasanya digunakan pada data dalam jumlah besar. Metode non hirarki yang sering digunakan adalah metode *K-means* (Johnson dan Wichern, 1988).

Menurut Sharma (1996), tujuan utama metode ini adalah melakukan pengelompokan berdasarkan kriteria tertentu sehingga objek-objek tersebut mempunyai ragam dalam kelompok (within cluster) relative lebih kecil dibandingkan ragam antar kelompok

(between cluster). Sedangkan menurut Johnson dan Winchern (1988), tujuan utama dari analisis cluster (kelompok) adalah ingin menentukan adanya kelompok-kelompok dari data yang tersedia. Ada beberapa manfaat dari analisis kelompok ini yaitu antara lain mulai dari sekedar eksplorasi data dengan cara melihat adanya pengelompokan, mereduksi dimensi data dan dapat pula sebagai acuan stratifikasi dalam teknik sampling.

Hasil pengelompokkan bergantung pada beberapa hal antara lain kriteria obyek yang diteliti, variabel yang diamati (skala pengukuran), ukuran ketidakmiripan sampai dengan teknik pengelompokannya. secara umum Analisis *Cluster* dibagi menjadi dua yaitu metode hirarki dan non hirarki. Metode hirarki yang terkenal antara lain *single linkage*, *complete linkage* dan *average linkage* disamping masih banyak metode lainnya. Sedangkan metode non hirarki yang digunakan adalah metode *K-means*.

#### 2.3.1 Metode K-Means

K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster atau kelompok. Metode ini mempartisi data ke dalam cluster atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan dari data clustering adalah untuk meminimalkan objective function yang diset dalam proses clustering, yang pada umumnya berusaha meminimalkan variasi di dalam suatu cluster dan memaksimalkan variasi antarcluster (Agusta, 2007).

Penggunaan pengelompokan non-hirarki mengasumsikan banyaknya kelompok diketahui dan telah didefinisikan sebelumnya (Dillon dan Goldstein, 1984). Menurut Yeung dan Ruzzo (2001), banyaknya kelompok merupakan input dari *k-means clustering. Cluster* dideskripsikan oleh centroid, yang merupakan pusat *cluster*. Tiap-tiap objek ditandai pada *centroid* dengan jarak Euclid terdekat. Banyaknya kelompok (*k*), diperoleh dengan menerapkan beberapa nilai *k* yang berbeda dan memilih solusi nilai *k* yang paling sesuai (Lattin *et al*, 2006).

Metode *K-Means* adalah metode *cluster* non hirarki yang popular (Afifi dan Clark, 1990). Menurut Agusta (2007), langkah-langkah algoritma *K-Means* adalah sebagai berikut :

- 1. Menyusun partisi awal derajat keanggotaan ke dalam setiap kelompok secara acak. Derajat keanggotaan mempunyai nilai 0 atau 1. Derajat keanggotaan kelompok merupakan peluang suatu objek data menjadi bagian dari kelompok tertentu. Derajat keanggotaan kelompok untuk setiap objek data pada awal analisis dapat dilakukan dengan membangkitkan bilangan acak sebagai elemen matriks partisi awal. Apabila suatu objek data merupakan anggota suatu kelompok, maka derajat keanggotaan bernilai 1 dan sebaliknya.
- 2. Menghitung nilai pusat kelompok dengan indeks *i* menunjukkan objek, *j* menunjukkan variabel dan *k* menunjukkan kelompok berdasarkan persamaan :

$$v_{jk} = \frac{\sum_{i=1}^{n} (x_{ij}\mu_{jk})}{\sum_{i=1}^{n} \mu_{ik}}$$
 (2.14)

$$i = 1, 2, ..., n; j = 1, 2, ..., m; k = 1, 2, ..., c$$

di mana:

 $v_{ik}$  = nilai pusat kelompok peubah ke-j kelompok ke-k

m =banyaknya peubah

c = banyaknya kelompok

xij = objek data ke-i pada peubah ke-j

 $\mu_{ik}$  = derajat keanggotaan objek data ke-i pada kelompok ke-k

3. Menghitung fungsi objektif berdasarkan persamaan:

$$J = \sum_{i=1}^{n} \sum_{k=1}^{c} \left( \left[ \sum_{j=1}^{m} (x_{ij} - v_{jk})^{2} \right] \mu_{ik} \right)$$
 (2.15)

di mana J merupakan nilai fungsi objektif. Pada langkah ini dilihat apakah nilai fungsi objektif telah konvergen berdasarkan persamaan :

$$|J_i - J_{i-1}| \leq \xi$$

di mana :  $\xi$  = kesalahan terkecil yang diharapkan i = langkah iterasi

Jika masih belum konvergen maka dilakukan penentuan derajat keanggotaan yang baru berdasarkan persamaan :

$$\mu_{ik} = \sqrt{\sum_{j=1}^{m} (x_{ij} - v_{jk})^2}$$
 (2.16)

di mana jarak terpendek pada setiap baris diberi nilai 1 sedangkan yang lain diberi nilai 0.

4. Jika nilai fungsi objektif telah konvergen, pengelompokan dari derajat keanggotaan setiap objek data dalam suatu kelompok ditentukan berdasarkan jarak terpendek hasil perhitungan jarak Euclid. Jarak terpendek pada setiap baris diberi nilai 1 sedangkan yg lain diberi nilai 0. Nilai 1 menunujukkan bahwa objek data menjadi anggota dalam suatu kelompok.

#### 2.4 Indeks Validitas

Setelah melakukan analisis *cluster*, perlu dilakukan validitas *cluster* karena penentuan jumlah kelompok dalam analisis *cluster* belum ada dasar yang kuat mengenai jumlah kelompok terbaik. Validitas cluster merupakan prosedur untuk mengevaluasi hasil analisis *cluster* secara kuantitatif sehingga dihasilkan kelompok optimum. Kelompok optimum yaitu kelompok yang mempunyai jarak yang padat antar individu dalam kelompok dan terisolasi dari kelompok lain dengan baik (Dubes dan Jain, 1988).

Menurut Kenesei et al. (2004), jumah *cluster* yang didapatkan bisa saja salah atau cluster yang didapatkan belum mewakili pengelompokan dari keseluruhan data. Terdapat dua cara yang dapat dilakukan untuk medapatkan jumlah cluster yang tepat dari data:

- Memulai pengelompokan data dengan k yang cukup besar dan mereduksi cluster yang memiliki kesamaan dengan menggunakan suatu kriteria tertentu. Cara ini disebut compatible cluster merging
- Menggunakan indeks validitas cluster untuk mendapatkan pengelompokkan yang terbaik.

### 2.4.1 Gap Statistic

Metode ini pada mulanya diperkenalkan oleh Tibshirani, *et al.* (2001) metode *Gap Statistic* dapat digunakan dalam menentukan jumlah kelompok optimum pada suatu himpunan data dengan

membangkitkan data referensi distribusi uniform. Ide dari *gap statistic* secara umum adalah mengusulkan dispersi dalam kelompok versus jumlah kelompok.

Gap statistic merupakan metode untuk menduga kelompok optimum pada analisis *cluster*. Teknik ini berdasar pada perubahan dispersi dalam *cluster* dengan peningkatan jumlah kelompok dari data (Arima et al, 2008). Secara umum dengan menaikkan jumlah kelompok, akan menurunkan ukuran kesalahan atau dispersi dalam kelompok (Kariyam dan Subanar, 2002).

Secara detil *Gap statistic* dapat dijelaskan sebagai berikut : Misalkan  $\{Xij\}$  dengan i=1,2,...,n dan j=1,2,...,p terdiri dari himpunan data dengan p (peubah) pada observasi *independent* dengan p objek. Kemudian data dikelompokkan menjadi p kelompok yaitu p dengan p dengan p menunjukkan pengamatan pada kelompok ke-p. Kemudian didefinisikan sebagai berikut :

$$Dr = \sum_{x_i, x_j} d(x_i, x_j)$$
 (2.17)

di mana:

Dr = jarak Euclid data observasi

 $d(x_i, x_j)$  = jarak antara objek ke-*i* dan objek ke-*j* di mana  $i \neq j$ 

Kemudian menghitung  $W_k$ :

$$W_k = \sum_{r=1}^k \frac{1}{2n_r} Dr (2.18)$$

dengan  $W_k$  adalah jumlah kuadrat dalam *cluster*, sedangkan nr adalah banyaknya observasi (anggota) kelompok ke-r.

Prosedur Gap Statistic dapat dijelaskan sebagai berikut:

- 1. Mengelompokkan data dan mengubah-ubah banyaknya kelompok mulai dari k = 1, 2, ...,n, yang masing-masing memberikan ukuran dispersi dalam kelompok  $W_k$ , dengan k = 1, 2, ...,n
- 2. Menentukan *k* optimum metode *Gap Statistic* dengan langkahlangkah:
  - a. Mendapatkan nilai  $W_k$  untuk tiap-tiap k (banyak cluster) kemudian menghitung Log  $W_k$

- Melakukan resampling (dari data simulasi) dengan pengembalian sebanyak B kali dengan distribusi referensi uniform
- c. Mendapatkan nilai  $W^*_{kb}$  dari poin (b) dimana b = 1, 2, ..., B dan k = 1, 2, ..., K
- d. Menghitung kelompok optimum dengan Gap Statistic:

$$Gap(k) = \left[\frac{1}{B}\right] \sum_{b} \log(\mathbf{W}^*_{kb}) - \log(W_k)$$
 (2.19)

di mana B merupakan resampling dengan distribusi uniform dan B = n (jumlah observasi)

e. Menghitung standar deviasi (sd<sub>k</sub>):

$$sd_k = \left[ \left[ \frac{1}{B} \right] \sum_{b} \{ \log(W^*_{kb}) - l \}^2 \right]^{1/2}$$
 (2.20)

Di mana:

$$l = \left[\frac{1}{B}\right] \sum_{b} \log(\mathbf{W}^*_{kb}) \tag{2.21}$$

f. Menghitung  $S_k$  dengan rumus:

$$s_k = sd_k \sqrt{(1 + \frac{1}{B})}$$
 (2.22)

g. Menghitung nilai diffu dengan persamaan:

$$diffu = Gap(k+1) - s_{k+1}$$
 (2.23)

h. Jumlah kelompok (*k*) optimum diperoleh dengan membandingkan nilai *Gap* (*k*) dengan nilai *diffu*. Sedemikian hingga nilai optimum diperoleh dari :

$$Gap(k) \ge diffu$$
 (2.24)

(Tibshirani et al, 2001)

#### 2.4.2 Silhouette

Rosseeuw (1987) mengajukan ukuran statistik yang disebut dengan silhouette untuk menyeleksi jumlah kelompok yang optimal. Kelebihan silhouette adalah dapat ditampilkan dalam bentuk grafik dua dimensi.

Beberapa tahapan untuk prosedur silhouette yaitu:

1. Untuk tiap-tiap objek *i* yang berada di kelompok *A*, dihitung :

$$a(i) = \frac{1}{n_4 - 1} \sum_{j \neq k, j \neq i} d_{ij}$$
 (2.25)

Di mana

 $n_A$  = banyaknya observasi pada kelompok A

a(i) = rata-rata jarak obyek ke-I ke semua obyek yang lain dalam kelompok A (dimana  $i = 1, 2, ..., n_r$ )

2. Ada beberapa kelompok lain yang berbeda dengan A, kemudian dihitung:

$$d_{i,C} = \frac{1}{n_C} \sum_{j \neq C} d_{ij} \tag{2.26}$$

Di mana

n<sub>C</sub> = banyaknya observasi pada kelompok C

 $d_{i,C}$  = rata-rata jarak obyek ke-I ke seluruh obyek di kelompok C (lainnya)

3. Setelah menghitung  $d_{i,C}$  untuk seluruh kelompok  $C \neq A$ , diambil nilai terkecil

$$b(i) = \min_{c \neq k} d_{ik} \tag{2.27}$$

Minimum dari rata-rata obyek ke-i ke semua obyek kelompok yang lain. Missal, kelompok B mencapai minimum dilihat dari  $d_{i,B} = b(i)$  maka dinamakan neighbour dari obyek ke-i. ini adalah second-best cluster dari obyek ke-i

4. Menghitung statistik silhouette yang didefinisikan:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}.$$
 (2.28)

Di mana:

 $i = 1, 2, \ldots, nr$ 

Nr = banyaknya observasi pada kelompok ke-r

Nilai s(i) terletak antara -1 sampai dengan 1, dengan interpretasi sebagai berikut :

 $s(i) \approx 1$  maka obyek ke-i terkelompok dengan baik (di A)

 $s(i) \approx 0$  maka obyek ke-*i* terletak diantara dua kelompok (A dan B)

 $s(i) \approx -1$  maka obyek ke-*i* tidak terkelompok dengan baik (tidak di *A* dan *B*)

Jumlah kelompok (k) optimum diestimasi dari harga k yang paling memaksimumkan nilai rata-rata s(i).

#### 2.5 Ukuran Kedekatan

Konsep dasar pengukuran pada analisis kelompok adalah pengukuran jarak (distance) dan kesamaan (similarity). Distance adalah ukuran jarak pisah antar objek sedangkan similarity adalah ukuran kedekatan. Pengukuran jarak (distance type measure) digunakan untuk data yang bersifat metrik (Dillon dan Goldstein, 1984).

Pengelompokan dimaksudkan untuk membentuk kelompokkelompok variabel sedemikian sehingga keragaman variabel dalam cluster haruslah lebih kecil daripada keragaman antar cluster. Untuk membentuk anggota dalam suatu cluster didasarkan pada ukuran similarity (kemiripan) dan dissimilarity (ketidakmiripan) masingmasing variabel.

Similarity adalah ukuran kemiripan yang digunakan untuk menemukan kemiripan antar objek. Misalkan s(i,j) adalah koefisien kemiripan, jika objek i dan j mirip maka s(i,j) memiliki nilai yang besar. Dan sebaliknya, , jika objek i dan j tidak mirip maka s(i,j) memiliki nilai yang kecil. Untuk semua objek i dan j, ukuran kemiripan harus mengikuti kondisi sebagai berikut :

a. 
$$0 \le s(i,j) \le 1$$
  
b.  $s(i,i) = 1$   
c.  $s(i,j) = s(j,i)$ 

Dissimilarity atau ukuran ketidakmiripan digunakan untuk menemukan ketidakmiripan antar objek. Misalkan dis(i,j) adalah

koefisien ketidakmiripan, maka dis(i,j) bernilai kecil jika i dan j mirip. Dan sebaliknya, jika objek i dan j tidak mirip maka dis(i,j) bernilai besar. Seperti ukuran kemiripan, ukuran ketidakmiripan harus mengikuti kondisi sebagai berikut :

- a.  $0 \le dis(i,j) \le 1$
- b. dis(i,i) = 1
- c. dis(i,j) = dis(j,i)

(Toledo, 2001)

Ukuran kedekatan jarak merupakan ukuran yang paling sering digunakan, diterapkan untuk data berskala metrik. Ukuran kedekatan jarak yang sering digunakan pada analisis kelompok adalah ukuran jarak Euclid. Jarak Euclid adalah metrika yang paling sering digunakan untuk menghitung kesamaan dua vektor. Jarak Euclid menghitung akar dari kuadrat perbedaan dua vektor. Jika data dinyatakan dalam bentuk matriks X yang anggotanya  $X_{ij}$ , maka jarak Euclid didefinisikan sebagai berikut:

$$d(i,j) = \sqrt{\sum_{k=1}^{p} (X_{ik} - X_{jk})^{2}} \quad , i \neq j \quad (2.29)$$

di mana:

d(i,j): jarak Euclid objek data ke-i dan objek data ke-j pada

peubah ke-k

p : banyaknya peubah

 $X_{ik}$ : data ke-i pada peubah ke-k $X_{ik}$ : data ke-i pada peubah ke-k

Jarak Euclid digunakan bila peubah-peubah yang digunakan tidak berkorelasi dan memiliki satuan yang sama. Jika terjadi korelasi antarpeubah maka dilakukan transformasi terhadap data awal dengan menggunakan Analisis Komponen Utama(AKU). Meskipun demikian, Hartigan dalam Handayani (2002) mengatakan bahwa jarak Euclid dengan atau tanpa transformasi komponen utama akan sama, jika seluruh komponen utama digunakan.

# 2.6 Cluster Tightness Measure (CTM)

Cluster Tightness Measure (CTM) merupakan ukuran kebaikan dari hasil pengelompokan berdasarkan simpangan baku setiap peubah pada masing-masing kelompok, yang dirumuskan sebagai bertikut:

$$CTM = \frac{1}{k} \sum_{k=1}^{K} \left( \frac{1}{p} \sum_{j=1}^{p} \frac{{}_{k} \sigma_{j}}{{}_{n} \sigma_{j}} \right)$$
 (2.30)

di mana:

 $_{k}\sigma_{j}$ : simpangan baku pada kelompok ke-k untuk peubah ke-

 $\sigma_i$ : simpangan baku seluruh data untuk peubah ke-j

p: banyaknya peubah

K: banyaknya kelompok

Kelompok yang terbentuk dikatakan baik jika memiliki nilai CTM terkecil.

(Epps dan Ambikairajah, 2008)

### 2.7 Sumber Daya Manusia

Menurut Winarni (2010), Sumber Daya Manusia (SDM) harus didefinisikan bukan dengan apa yang sumber daya manusia lakukan, tetapi apa yang sumber daya manusia hasilkan. Sumber Daya Manusia yang berkualitas merupakan aset yang penting bagi pembangunan di berbagai aspek kehidupan masyarakat. SDM yang berkualitas dapat tercipta apabila setiap warga masyarakat dapat memperoleh pelayanan pendidikan dan kesehatan yang memadai, merata dan bermutu. Dengan modal SDM yang berkualitas, diharapkan Bangsa Indonesia dapat berdiri sejajar dengan bangsa-bangsa lain di dunia, apalagi di era globalisasi yang ditandai dengan persaingan internasional yang semakin ketat (Badan Pusat Statistik, 2008).

## 2.7.1 Indeks Pembangunan Manusia

Pembangunan manusia adalah perubahan yang terjadi pada manusia dilihat dari sisi ekonomi dan sosial. Indeks Pembangunan Manusia (IPM) merupakan ukuran capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup. IPM menggambarkan beberapa komponen, yaitu capaian umur panjang dan sehat yang mewakili bidang kesehatan; angka melek huruf, partisipasi sekolah dan rata-rata lamanya bersekolah mengukur kinerja pembangunan bidang pendidikan; dan kemampuan daya beli masyarakat terhadap sejumlah kebutuhan pokok yang dilihat dari rata-rata besarnya pengeluaran per kapita.

Pembangunan manusia yang berhasil akan membuat usia ratarata masyarakat meningkat dengan peningkatan pengetahuan yang bermuara pada peningkatan kualitas SDM. Pencapaian hal tersebut selanjutnya akan meningkatkan produktivitas sehingga akhirnya akan meningkatkan mutu hidup dalam arti hidup yang layak.

(Badan Pusat Statistik, 2008)

Konsep Pembangunan Manusia yang dikembangkan oleh Perserikatan Bangsa-Bangsa (PBB), menetapkan peringkat kinerja pembangunan manusia pada skala 0,0 – 100,0 dengan katagori sebagai berikut :

□Tinggi: IPM lebih dari 80,0

□ Menengah Atas : IPM antara 66,0 – 79,9

☐ Menengah Bawah : IPM antara 50,0 – 65,9

Rendah: IPM kurang dari 50,0

(Sukmaraga, 2011)

### 2.7.2 Angka Harapan Hidup

Kemampuan untuk bertahan hidup lebih lama diukur dengan indikator harapan hidup pada saat lahir (*life expectacy at birth*). Perhitungan angka harapan hidup pada saat lahir dihitung berdasarkan data rata-rata jumlah anak lahir hidup dan rata-rata jumlah anak masih hidup menurut kelompok umur ibu 15-49 tahun, yang bersumber dari data hasil Susenas dengan memperhatikan trend hasil Sensus Penduduk. Indeks yang digunakan untuk mengukur angka harapan hidup disebut Indeks Harapan Hidup (IHH).

### 2.7.3 Tingkat Pendidikan

Komponen tingkat pendidikan diukur dengan Indeks Pendidikan (IP) yang diperoleh dari dua indikator, yaitu angka melek huruf dan rata-rata lama sekolah. Angka melek huruf adalah persentase dari penduduk usia 15 tahun ke atas yang bisa membaca dan menulis huruf latin atau huruf lainnya, terhadap jumlah penduduk usia 15 tahun atau lebih. Indikator ini diberi bobot dua per tiga. Bobot sepertiga sisanya diberikan pada indicator rata-rata lama bersekolah (Mean Year Schooling/YMS), vaitu rata-rata jumlah tahun yang telah dihabiskan oleh penduduk usia 15 tahun ke atas di seluruh jenjang pendidikan formal yang pernah dijalani. Indicator ini dihitung dari variabel pendidikan tertinggi yang ditamatkan dan tingkat pendidikan yang sedang diduduki.

# 2.7.4 Standar Layak Hidup

Standar layak hidup atau *Purchasing Power Parity* (PPP) merupakan komponen yang menggunakan Produk Domestik Bruto (PDB) riil perkapita yang disesuaikan untuk mengukur standar hidup layak menggunakan rata-rata pengeluaran perkapita riil yang disesuaikan dengan formula Atkinson. PDB riil adalah Produk Domestik Bruto atas dasar harga konstan yang dihitung dengan memasukkan pengaruh dari harga. PDB nominal merupakan merupakan PDB atas dasar harga berlaku. Jadi, PDB riil melengkapi atau mengoreksi angka yang terdapat pada PDB nominal. Formula Atkinson merupakan formula yang digunakan untuk menentukan PPP dari nilai riil pengeluaran perkapita.

Penentuan standar harga untuk menghitung PPP (daya beli) di wilayah Jawa Timur, standar harga yang digunakan adalah Kota Malang. Kontribusi yang besar dari angka PPP dapat tercapai seiring dengan peningkatan kesejahteraan penduduk sebagai dampak dari pertumbuhan ekonomi dan pemerataan pendapatan.



### BAB III METODE PENELITIAN

### 3.1 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data Penyusunan Kinerja Makro Ekonomi dan Sosial Jawa Timur tentang Pembangunan Manusia dari Tahun 2005 sampai Tahun 2008 yang diperoleh dari Badan Pusat Statistik Provinsi Jawa Timur. Data penelitian ini terdiri atas 4 variabel penelitian yaitu Indeks Harapan Hidup (IHH), Indeks Pendidikan (IP), Indeks *Purchasing Power Parity* (PPP) dan Indeks Pembangunan Manusia (IPM) pada sampel berukuran 38. Data yang digunakan adalah data dalam bentuk persentase. Data penelitian secara lengkap dapat dilihat pada Lampiran 1.

#### 3.2 Metode Penelitian

### 3.2.1 Tahap Persiapan

Tahap persiapan dilakukan untuk menguji korelasi data. Data penelitian terlebih dahulu diuji korelasinya dengan menggunakan uji korelasi Pearson, karena data yang digunakan adalah data parametrik dan ukuran kedekatan yang digunakan adalah jarak Euclid. Jika terjadi korelasi antarpeubah, maka dilakukan transformasi terhadap data awal dengan menggunakan Analisis Komponen Utama(AKU). Langkah-langkah tahap persiapan adalah sebagai berikut:

- 1. Melakukan uji korelasi sebagai berikut:
  - a. Menghitung nilai koefisien korelasi (r) menggunakan persamaan (2.1)
  - b. Menentukan taraf signifikansi (α)
  - c. Menghitung nilai t dengan persamaan (2.2)
  - d. Mencari nilai ttabel dengan derajat bebas = n-2
  - e. Menghitung *p-value* dengan persamaan (2.3)
  - f. Membandingkan *p-value* dengan nilai α
- 2. Melakukan analisis komponen utama sebagai berikut :
  - a. Menggunakan matriks ragam peragam sebagai masukan
  - b. Menghitung eigen value dan eigen vektor untuk digunakan dalam analisis kelompok
  - c. Menghitung skor komponen utama

- d. Melakukan pemilihan komponen utama untuk digunakan dalam analisis kelompok
- e. Menggunakan *k* komponen utama yang memiliki proporsi keragaman total lebih dari 90 %, untuk digunakan dalam analisis kelompok

## 3.2.2 Tahap Analisis Data

Langkah-langkah analisis pada penelitian ini adalah sebagai berikut:

- 1. Melaksanakan prosedur pengelompokan dengan menggunakan metode *K-means*, dengan jumlah kelompok adalah k dan besar k = 3; k = 4; dan k = 5.
- 2. Mengestimasi *k* optimum metode *Gap Statistic* dengan langkah-langkah:
  - a. Mendapatkan nilai  $W_k$  untuk tiap-tiap k (banyak cluster) kemudian menghitung Log  $W_k$ .
  - b. Melakukan resampling dengan pengembalian sebanyak *B* kali dengan distribusi referensi uniform
  - c. Mendapatkan nilai  $W^*_{kb}$  dari poin (b) dimana b = 1, 2, ..., B dan k = 1, 2, ..., K
  - d. Menghitung kelompok optimum dengan *Gap Statistik* menggunakan persamaan (2.19)
  - e. Menghitung standar deviasi  $(sd_k)$  menggunakan persamaan (2.20)
  - f. Menghitung  $s_k$  menggunakan persamaan (2.22)
  - g. Menghitung nilai *diffu* menggunakan persamaan (2.23)
  - h. Mengestimasi jumlah kelompok (k) optimum menggunakan persamaan (2.24)
- 3. Mengestimasi k optimum metode *Silhouette* dengan langkahlangkah:
  - a. Mencari besarnya *a(i)*, yaitu rata-rata jarak obyek ke-i ke semua obyek lain dalam kelompoknya sendiri menggunakan persamaan (2.25)
  - b. Mencari besarnya b(i), yaitu rata-rata jarak obyek ke-i ke semua obyek lain diluar kelompoknya yang paling minimum menggunakan persamaan (2.27)
  - c. Menghitung statistik silhouette menggunakan persamaan (2.28)

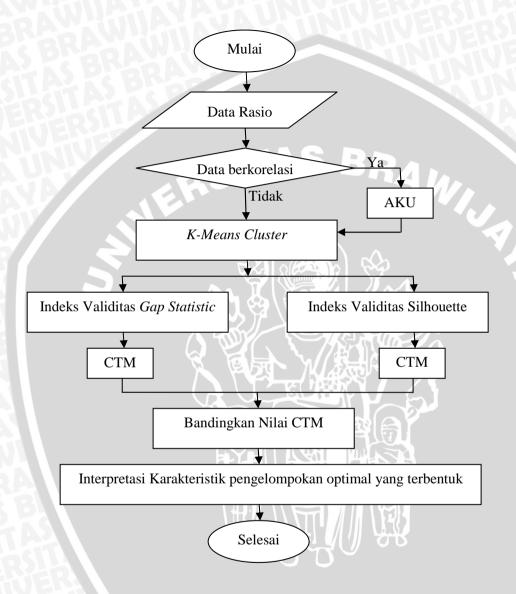
- d. Jumlah kelompok (k) optimum diestimasi dengan cara mencari nilai rata-rata s(i) yang paling maksimal.
- 4. Menghitung nilai CTM dari masing-masing indeks validitas menggunakan persamaan (2.30)
- 5. Membandingkan nilai CTM untuk kedua indek validitas, indeks validitas terbaik ditunjukkan oleh nilai CTM minimum.
- 6. Membandingkan perubahan kelompok data Penyusunan Kinerja Makro ekonomi dan Sosial Jawa Timur tentang Pembangunan Manusia dari tahun ke tahun.

Analisis dan perhitungan pada penelitian ini menggunakan software Minitab 14, software R 2.13.0, dan Microsoft Excel 2007. Software Minitab 14 digunakan untuk uji koefisien korelasi dan Analisis Komponen Utama, software R 2.13.0 digunakan untuk analisis kelompok dan menghitung indeks validitas, dan Microsoft Excel 2007 digunakan untuk menghitung nilai CTM.

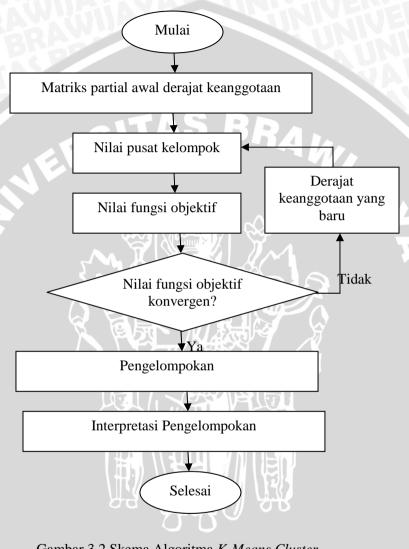
### 3.3 Diagram Alir

Tahapan-tahapan pada penelitian ini dapat dilihat pada diagram alir pada Gambar 3.1.





Gambar 3.1 Diagram Alir Metode Penelitian



Gambar 3.2 Skema Algoritma K-Means Cluster



### BAB IV HASIL DAN PEMBAHASAN

### 4.1 Uji Koefisien Korelasi

Analisis *cluster* dengan menggunakan ukuran kedekatan jarak Euclid mensyaratkan tidak terdapat korelasi antar variabel. Adanya korelasi antar variabel sangat berpengaruh terhadap hasil pengelompokan.

Data yang dianalisis adalah empat data tentang Penyusunan Kinerja Makro ekonomi dan Sosial Jawa Timur tentang Pembangunan Manusia, yang selanjutnya keempat data tersebut diberi nama Data 1 untuk Data Tahun 2005, Data 2 untuk Data Tahun 2006, Data 3 untuk Data Tahun 2007 dan Data 4 untuk Data Tahun 2008. Hal ini dilakukan untuk mempermudah interpretasi.

Setelah dilakukan uji koefisien korelasi terhadap Data 1, Data 2, Data 3 dan Data 4, hasil pengujian koefisien korelasi secara lengkap dapat dilihat pada Lampiran 2.

Tabel 4.1. Hasil Pengujian Koefisien Korelasi Data 1

		1 4 4 7 1 2 2 3 7 4 6	•	
		IHH	IP (	PPP
IP	Pearson correlation	0.779		
11	P-Value	0.00	भक्ति।	
PPP	Pearson correlation	0.730	0.852	
	P-Value	0.00	0.00	
IPM	Pearson correlation	0.907	0.967	0.886
	P-Value	0.00	0.00	0.00

Tabel 4.1 menunjukkan nilai p sebesar 0.000 untuk masing-masing variabel. Nilai p dari hasil pengujian koefisien korelasi Data 1 lebih kecil daripada nilai resiko berbuat salah ( $\alpha=0.05$ ). Hingga diputuskan untuk menolak  $H_0$  yang berarti terdapat korelasi antar variabel tersebut. Adanya korelasi antar variabel juga dapat dilihat dari besarnya nilai koefisien korelasi antar masing-masing variabel yang nilainya mendekati 1. Sebagai contoh nilai koefisien korelasi antara variabel IPM dengan variabel IHH adalah sebesar 0.907, nilai

koefisien korelasi antara variabel IPM dengan variabel IP adalah sebesar 0.967 dan seterusnya sebagaimana pada Tabel 4.1.

Tabel 4.2. Hasil Pengujian Koefisien Korelasi Data 2

401		IHH	IP	PPP
IP	Pearson correlation	0.775		
11	P-Value	0.00		
PPP	Pearson correlation	0.680	0.802	Ala
111	P-Value	0.00	0.00	
IPM	Pearson correlation	0.904	0.966	0.844
II WI	P-Value	0.00	0.00	0.00

Tabel 4.2 menunjukkan nilai p sebesar 0.000 untuk masing-masing variabel. Nilai p dari hasil pengujian koefisien korelasi Data 2 lebih kecil daripada nilai resiko berbuat salah ( $\alpha=0.05$ ). Hingga diputuskan untuk menolak  $H_0$  yang berarti terdapat korelasi antar variabel tersebut. Adanya korelasi antar variabel juga dapat dilihat dari besarnya nilai koefisien korelasi antar masing-masing variabel yang nilainya mendekati 1. Sebagai contoh nilai koefisien korelasi antara variabel IPM dengan variabel IHH adalah sebesar 0.904, nilai koefisien korelasi antara variabel IPM dengan variabel IPM dengan variabel IP adalah sebesar 0.966 dan seterusnya sebagaimana pada Tabel 4.2.

Tabel 4.3. Hasil Pengujian Koefisien Korelasi Data 3

		IHH	IP	PPP
IP	Pearson correlation	0.797		5)3
п	P-Value	0.00	7	8
PPP	Pearson correlation	0.757	0.835	
	P-Value	0.00	0.00	
IPM	Pearson correlation	0.915	0.969	0.881
	P-Value	0.00	0.00	0.00

Tabel 4.3 menunjukkan nilai p sebesar 0.000 untuk masing-masing variabel. Nilai p dari hasil pengujian koefisien korelasi Data 3 lebih

kecil daripada nilai resiko berbuat salah ( $\alpha=0.05$ ). Hingga diputuskan untuk menolak  $H_0$  yang berarti terdapat korelasi antar variabel tersebut. Adanya korelasi antar variabel juga dapat dilihat dari besarnya nilai koefisien korelasi antar masing-masing variabel yang nilainya mendekati 1. Sebagai contoh nilai koefisien korelasi antara variabel IPM dengan variabel IHH adalah sebesar 0.915, nilai koefisien korelasi antara variabel IPM dengan variabel IP adalah sebesar 0.969 dan seterusnya sebagaimana pada Tabel 4.3.

Tabel 4.4. Hasil Pengujian Koefisien Korelasi Data 4

		IHH	IP	PPP
IP	Pearson correlation	0.799		
	P-Value	0.00	rOn	
PPP	Pearson correlation	0.606	0.637	
	P-Value	0.00	0.00	
IPM	Pearson correlation	0.916	0.961	0.749
	P-Value	0.00	0.00	0.00

Tabel 4.4 menunjukkan nilai p sebesar 0.000 untuk masing-masing variabel. Nilai p dari hasil pengujian koefisien korelasi Data 4 lebih kecil daripada nilai resiko berbuat salah ( $\alpha=0.05$ ). Hingga diputuskan untuk menolak  $H_0$  yang berarti terdapat korelasi antar variabel tersebut. Adanya korelasi antar variabel juga dapat dilihat dari besarnya nilai koefisien korelasi antar masing-masing variabel yang nilainya mendekati 1. Sebagai contoh nilai koefisien korelasi antara variabel IPM dengan variabel IHH adalah sebesar 0.916, nilai koefisien korelasi antara variabel IPM dengan variabel IP adalah sebesar 0.961 dan seterusnya sebagaimana pada Tabel 4.4.

Hasil uji hipotesis tersebut menunjukkan bahwa perlu dilakukan analisis komponen utama terhadap keempat data penelitian, dengan tujuan menghilangkan korelasi antar variabel.

### 4.2 Analisis Komponen Utama

Analisis komponen utama mentransformasi sekumpulan variabel menjadi variabel baru di mana variabel yang dihasilkan tidak saling berkorelasi dan merupakan kombinasi linier variabel asal. Pada keempat data penelitian, perlu dilakukan analisis komponen utama karena terdapat korelasi antar variabel dari masing-masing data.

Hasil analisis komponen utama yaitu skor komponen utama, nilai eigen, dan persentase keragaman disajikan secara lengkap pada Lampiran 3. Analisis komponen utama akan menghasilkan skor komponen utama, di mana skor komponen utama yang diperoleh akan digunakan sebagai masukan dalam analisis *cluster*.

Skor komponen utama yang digunakan sebagai masukan dalam analisis *cluster* adalah *k* komponen utama yang memiliki proporsi keragaman total lebih dari 90%. Jadi, setelah didapatkan skor komponen utama harus dilakukan pemilihan komponen utama yang akan digunakan sebagai masukan dalam analisis kelompok.

Hasil analisis komponen utama dari keempat data tersebut adalah sebagai berikut :

### 4.2.1 Hasil Analisis Komponen Utama Data 1

Berdasarkan hasil analisis komponen utama Data 1, diperoleh nilai eigen komponen utama pertama sebesar 3.5660, nilai eigen komponen utama kedua sebesar 0.2847, nilai eigen komponen utama ketiga sebesar 0.1493 dan nilai eigen komponen utama keempat sebesar 0.000. Dari nilai eigen yang diperoleh dapat dilihat bahwa sebenarnya hanya komponen utama pertama yang memegang peranan penting terhadap keragaman data asal. Karena, hanya komponen utama pertama yang memiliki nilai eigen lebih dari 1, dapat dilihat dari nilai eigen sebesar 3.566. Namun, komponen tersebut hanya dapat menjelaskan keragaman total sebesar 89.2% dari keragaman data semula, sehingga masih banyak informasi data awal yang hilang jika digunakan patokan tersebut sebagai masukan dalam analisis *cluster*.

Nilai keragaman total untuk komponen utama pertama adalah sebesar 0.892 yang artinya komponen utama pertama mampu menjelaskan keragaman total sebesar 89.2% dari keragaman data semula. Sedangkan, apabila digunakan dua komponen utama pertama didapatkan nilai keragaman total sebesar 0.963, yang berarti dua

komponen utama pertama mampu menjelaskan keragaman total sebesar 96.3% dari keragaman data semula.

Jadi, jika digunakan dua komponen utama pertama sebagai masukan dalam analisis *cluster*, maka kedua komponen utama tersebut dapat menggantikan variabel-variabel asli tanpa kehilangan banyak informasi. Skor komponen utama, nilai eigen, dan persentase keragaman total dari hasil analisis komponen utama disajikan secara lengkap pada Lampiran 3.

### 4.2.2 Hasil Analisis Komponen Utama Data 2

Berdasarkan pendekatan akar karakteristik lebih besar dari satu, dari hasil analisis komponen utama Data 2, diperoleh nilai eigen komponen utama pertama sebesar 3.4931, nilai eigen komponen utama kedua sebesar 0.3262, nilai eigen komponen utama ketiga sebesar 0.1804 dan nilai eigen komponen utama keempat sebesar 0.000.

hanya satu komponen utama yang memegang peranan penting terhadap keragaman data asal. Yaitu komponen utama pertama. Karena, hanya komponen utama pertama yang memiliki nilai eigen yang lebih dari 1. Tetapi, komponen utama pertama tersebut hanya dapat menjelaskan keragaman total sebesar 87.3% dari keragaman data semula.

Nilai keragaman total untuk komponen utama pertama adalah sebesar 0.873 yang artinya komponen utama pertama mampu menjelaskan keragaman total sebesar 87.3% dari keragaman data semula. Sedangkan, apabila digunakan dua komponen utama pertama didapatkan nilai keragaman total sebesar 0.955, yang berarti dua komponen utama pertama mampu menjelaskan keragaman total sebesar 95.5% dari keragaman data semula.

Jadi, jika digunakan dua komponen utama pertama sebagai masukan dalam analisis *cluster*, maka kedua komponen utama tersebut dapat menggantikan variabel-variabel asli tanpa kehilangan lebih banyak informasi. Skor komponen utama, nilai eigen, dan persentase keragaman total dari hasil analisis komponen utama disajikan secara lengkap pada Lampiran 3.

### 4.2.3 Hasil Analisis Komponen Utama Data 3

Berdasarkan hasil analisis komponen utama Data 3, diperoleh nilai eigen komponen utama pertama sebesar 3.5812, nilai eigen komponen utama kedua sebesar 0.2506, nilai eigen komponen utama ketiga sebesar 0.1682 dan nilai eigen komponen utama keempat sebesar 0.000. Dari nilai eigen yang diperoleh, dapat dilihat bahwa sebenarnya hanya komponen utama pertama yang memegang peranan penting terhadap keragaman data asal. Karena, hanya komponen utama pertama yang memiliki nilai eigen lebih dari 1. Hal ini dapat dilihat dari nilai eigen sebesar 3.5812. Namun, komponen tersebut hanya dapat menjelaskan keragaman total sebesar 89,5% dari keragaman data semula, sehingga masih banyak informasi data awal yang hilang jika digunakan patokan tersebut sebagai masukan dalam analisis *cluster*.

Nilai keragaman total untuk komponen utama pertama adalah sebesar 0.895 yang artinya komponen utama pertama mampu menjelaskan keragaman total sebesar 89,5% dari keragaman data semula. Sedangkan, apabila digunakan dua komponen utama pertama didapatkan nilai keragaman total sebesar 0.958, yang berarti dua komponen utama pertama mampu menjelaskan keragaman total sebesar 95.8% dari keragaman data semula.

Jadi, jika digunakan dua komponen utama pertama sebagai masukan dalam analisis *cluster*, maka kedua komponen utama tersebut dapat menggantikan variabel-variabel asli tanpa kehilangan banyak informasi. Skor komponen utama, nilai eigen, dan persentase keragaman total dari hasil analisis komponen utama disajikan secara lengkap pada Lampiran 3.

### 4.2.4 Hasil Analisis Komponen Utama Data 4

Berdasarkan hasil analisis komponen utama Data 4, diperoleh nilai eigen komponen utama pertama sebesar 3.3511, nilai eigen komponen utama kedua sebesar 0.4467, nilai eigen komponen utama ketiga sebesar 0.2022 dan nilai eigen komponen utama keempat sebesar 0.000. Dari nilai eigen yang diperoleh dapat dilihat bahwa sebenarnya hanya komponen utama pertama yang memegang peranan penting terhadap keragaman data asal. Karena, hanya komponen utama pertama yang memiliki nilai eigen lebih dari 1, dapat dilihat dari nilai eigen sebesar 3.3511. Namun, komponen tersebut hanya dapat menjelaskan keragaman total sebesar 83.8% dari keragaman data

semula, sehingga masih banyak informasi data awal yang hilang jika digunakan patokan tersebut sebagai masukan dalam analisis *cluster*.

Nilai keragaman total untuk komponen utama pertama adalah sebesar 0.838 yang artinya komponen utama pertama mampu menjelaskan keragaman total sebesar 83.8% dari keragaman data semula. Sedangkan, apabila digunakan dua komponen utama pertama didapatkan nilai keragaman total sebesar 0.949, yang berarti dua komponen utama pertama mampu menjelaskan keragaman total sebesar 94.9% dari keragaman data semula.

Jadi, jika digunakan dua komponen utama pertama sebagai masukan dalam analisis *cluster*, maka kedua komponen utama tersebut dapat menggantikan variabel-variabel asli tanpa kehilangan lebih banyak informasi. Skor komponen utama, nilai eigen, dan persentase keragaman total dari hasil analisis komponen utama disajikan secara lengkap pada Lampiran 3.

# 4.3. Analisis K-Means Cluster dengan Indeks Validitas Gap Statistic

Setelah dilakukan pemilihan komponen utama, skor komponen utama yang telah didapatkan digunakan sebagai masukan untuk menentukan pengelompokan. Dengan demikian, dari Data 1, Data 2, Data 3 dan Data 4 digunakan dua komponen utama pertama sebagai masukan dalam analisis *cluster*.

Dalam analisis *k-means cluster* untuk mengelompokkan objek, terlebih dahulu ditentukan jumlah *cluster* (*k*) yang akan digunakan, dan selanjutnya dihitung nilai indeks validitas *Gap Statistic* dari masing-masing jumlah *cluster* yang telah ditetapkan. Berdasarkan hasil pengelompokan dengan metode *k-means cluster* dengan *k*=3, *k*=4 dan *k*=5 menggunakan indeks validitas *Gap Statistic* tersebut, diperoleh anggota setiap *cluster*, nilai *Gap Statistic* dan nilai *diffu* sebagaimana pada Lampiran 4.

Setelah dilakukan analisis *k-means cluster* dan validitas *cluster* dengan menggunakan indeks validitas *Gap statistic* didapatkan hasil seperti pada Tabel 4.5.

Tabel 4.5. Nilai Indeks Validitas *Gap statistic* (gap)

Data	Jumlah Kelompok	Gap statistic	Nilai diffu
<b>X</b>	3	0.5686075	0.08382931
1	4	0.6782011	0.38055730
3804	5	0.5663874	0.08772923
	3	0.6002252	0.18205910
2	4	0.6436619	0.23968340
	5	0.5437670	0.11559490
	3	0.2968127	0.05085269
3	4	0.3469865	0.17276230
	5	0.3543581	0.19501340
	3	0.6095027	0.10983270
4	4	0.5171042	0.17605870
	5	0.5723530	0.17206200

Tabel 4.5 merupakan ringkasan hasil perhitungan nilai validitas *cluster* dari indeks validitas *Gap Statistic*. Hasil perhitungan validitas *cluster* dari indeks validitas tersebut menunjukkan bahwa masing-masing jumlah *cluster* menghasilkan nilai validitas yang tidak sama. Pada indeks validitas *Gap Statistic* jumlah *cluster* optimum ditunjukkan oleh nilai gap yang memenuhi persamaan 2.22 dengan demikian nilai *cluster* optimum diperoleh dari  $gap(k) \ge diffu(k+1)$ .

Data 1 mengindikasika bahwa jumlah optimum *cluster* yang diperoleh adalah 3 karena pada *cluster* ke-3 diperoleh nilai *gap*  $(0.5686075) \ge diffu$  (0.38055730). Pada Data 2 diperoleh nilai *gap*  $(0.6002252) \ge diffu$  (0.23968340) pada *cluster* ke-3, sehingga didapatkan nilai optimum *cluster* adalah 3 *cluster*. Demikian juga Data 3 dan Data 4 didapatkan nilai *cluster* optimum sama dengan 3. Hal ini ditunjukkan oleh nilai *gap statistic* dan nilai *diffu* untuk Data 3 dan Data 4 secara berturut-turut adalah  $(0.2968127) \ge diffu$  (0.17276230) dan  $(0.6095027) \ge diffu$  (0.17605870) untuk *cluster* ke-3.

Jadi. indeks validitas *Gap statistic* memberikan hasil yang sama pada Data 1, Data 2, Data 3 dan Data 4. Jumlah *cluster* optimum yang diperoleh menggunakan indeks validitas *Gap statistic* adalah 3 *cluster*.

### 4.4. Analisis K-Means Cluster dengan Indeks Validitas Silhouette

Setelah dilakukan pemilihan komponen utama, skor komponen utama yang telah didapatkan digunakan sebagai masukan untuk menentukan pengelompokan. Dengan demikian, dari Data 1, Data 2, Data 3 dan Data 4 digunakan dua komponen utama pertama sebagai masukan dalam analisis *cluster*.

Dalam analisis *k-means cluster* untuk mengelompokkan objek, terlebih dahulu ditentukan jumlah *cluster* (*k*) yang akan digunakan, dan selanjutnya dihitung nilai indeks validitas *Silhouette* dari masingmasing jumlah *cluster* yang telah ditetapkan. Berdasarkan hasil pengelompokan dengan metode *k-means cluster* dengan *k*=3, *k*=4 dan *k*=5 menggunakan indeks validitas *Silhouette*, diperoleh anggota setiap *cluster* dan nilai *Silhouette* sebagaimana pada Lampiran 5.

Setelah dilakukan analisis *k-means cluster* dan validitas *cluster* dengan menggunakan indeks validitas *Silhouette* didapatkan hasil seperti pada Tabel 4.6

Tabel 4.6. Nilai Indeks Validitas Silhouette (S)

Data	Jumlah Kelompok	Silhouette
	3	0.5063859
	4	0.4474859
1	5	0.3990262
	3	0.49143
	4	0.4536038
2	5	0.43086
	3	0.4676076
	4	0.4746141
3	5	0.4763737

Tabel 4.6. (Lanjutan)

Data	Jumlah Kelompok	Silhouette
	3	0.4351177
4	4	0.3952428
#17-	5	0.4455983

Tabel 4.6 merupakan ringkasan hasil perhitungan nilai validitas cluster dari indeks validitas *Silhouette*. Dari tabel tersebut, tampak bahwa pada indeks validitas *Silhouette* jumlah *cluster* optimum ditunjukkan oleh nilai S yang paling mendekati satu, atau dengan kata lain ditunjukkan oleh nilai S yang paling besar. Jadi, pada Data 1 didapatkan jumlah *cluster* optimum yang diperoleh adalah 3 karena memiliki nilai S yang tertinggi yaitu 0.5063859. Demikian juga pada Data 2, jumlah *cluster* optimum yang didapatkan berdasarkan indeks validitas *Silhouette* adalah 3 *cluster* karena diperoleh nilai S tertinggi yaitu 0.4914300 pada *cluster* ke-3.

Berbeda dengan Data 3, pada data ini diperoleh nilai optimum cluster adalah 5 cluster. Karena nilai S tertinggi adalah pada cluster ke-5 dengan nilai S adalah 0.4763737. Sedangkan pada Data 4 nilai cluster 0.4455983, yaitu nilai cluster tertinggi diperoleh pada cluster 5. Sehingga dapat disimpulkan bahwa nilai cluster optimum pada Data 4 adalah 5.

Jadi. indeks validitas *Silhouette* memberikan hasil yang tidak sama pada Data 1, Data 2, Data 3 dan Data 4. Jumlah *cluster* optimum yang diperoleh menggunakan indeks validitas *Silhouette* pada Data 1 dan Data 2 adalah 3 *cluster*. Sedangkan Jumlah *cluster* optimum yang diperoleh menggunakan indeks validitas *Silhouette* pada Data 3 dan Data 4 adalah 5 *cluster*.

# 4.5. Perbandingan Indeks Validitas *Gap Statistic* dan Indeks Validitas *Silhouette*

Jumlah *cluster* optimum yang diperoleh menggunakan kedua indeks validitas yaitu *Gap statistic* dan *Silhouette*, memberikan hasil yang tidak sama. Pada Data 1 dan Data 2 indeks validitas *Gap statistic* dan *Silhouette* memberikan hasil yang sama yaitu didapatkan jumlah *cluster* optimum adalah 3 cluater. Tetapi, pada Data 3 dan Data 4 menunjukkan hasil yang berbeda.

Berdasarkan perbedaan hasil kedua indeks validitas terbebut, diperlukan suatu ukuran kebaikan hasil pengelompokan. Dalam penelitian ini ukuran kebaikan hasil pengelompokan yang digunakan adalah *Cluster Tightness Measure* (CTM).

Cluster Tightness Measure merupakan ukuran kebaikan dari hasil pengelompokan berdasarkan simpangan baku setiap peubah pada masing-masing cluster. Nilai CTM untuk kedua indeks validitas tersebut disajikan pada Tabel 4.7.

Tabel 4.7. Nilai CTM Indeks Validitas Gap Statistic dan Silhouette

Data	Jumlah	Nilai C	TM
Data	Kelompok	Gap statistic	Silhouette
	3	1.00564	0.27502
1	4	1.00165	0.30725
	5	1.04531	0.37561
	314	1.02222	0.28541
2	47	0.99865	0.30067
	5	0.97846	0.34919
	3	0.98069	0.27906
3	4	0.97296	0.29085
	5	0.99219	0.31672
	3	0.99625	0.29937
4	4	0.94129	0.33353
	5	1.01556	0.39940

Tabel 4.7 merupakan ringkasan hasil perhitungan nilai CTM dari indeks validitas *Gap Statistic* dan indeks validitas *Silhouette*. Nilai

CTM yang terkecil menunjukan indeks validitas terbaik dalam penentuan jumlah optimum cluster untuk setiap data penelitian. Sebagai contoh pada Data 1, nilai CTM untuk indeks validitas Gap statistic dan Silhouette pada jumlah cluster (k) sama dengan 3 secara berturut-turut adalah 1.00564 dan 0.27502. Indeks validitas Silhouette memberikan niai CTM yang lebih kecil. Berdasarkan nilai CTM tersebut dapat disimpulkan bahwa indeks validitas Silhouette merupakan indeks validitas yang lebih baik dibandingkan indeks validitas Gap statistic untuk Data 1 dengan jumlah pengelompokan sama dengan 3. Selanjutnya untuk Data 1, nilai CTM untuk indeks validitas Gap statistic dan Silhouette pada jumlah cluster (k) sama dengan 4 secara berturut-turut adalah 1.00165 dan 0.30725. Indeks validitas Silhouette memberikan nilai yang lebih kecil, sehingga indeks validitas Silhouette merupakan indeks validitas yang lebih baik dibandingkan indeks validitas Gap statistic untuk Data 1 dengan jumlah pengelompokan sama dengan 4.

Nilai CTM yang diperoleh dari keseluruhan hasil pengelompokan, 100% nilai CTM dari indeks validitas *Silhouette* menunjukkan nilai yang lebih besar daripada nilai CTM yang dihasilkan indeks validitas *Gap statistic*. Jadi, hasil pengelompokan menunjukkan bahwa indeks validitas *Silhouette* merupakan indeks validitas yang lebih baik dalam penentuan jumlah optimum *cluster* dibandingkan indeks validitas *Gap statistic*.

Berdasarkan nilai CTM dapat disimpulkan indeks validitas Silhouette merupakan indeks validitas yang lebih baik dalam penentuan jumlah optimum cluster dibandingkan indeks validitas Gap Statistic pada analisis cluster metode K-means Cluster. Karena, Indeks validitas Silhouette menggunakan rata-rata jarak setiap titik ke titik lain dalam kelompok dan di luar kelompok. Sedangkan indeks validitas Gap statistic selalu menghasilkan nilai CTM yang lebih besar karena dari kelompok yang terbentuk cenderung menghasilkan ragam dalam kelompok yang lebih besar dibandingkan indeks validitas Silhouette. Menurut (Tibshirani et al., 2001) estimasi indeks validitas Gap statistic akan baik pada data yang terpisah dengan baik, ketika himpunan data tidak terpisah dengan baik maka estimasi jumlah kelompok tidak akan bisa dijelaskan secara tepat.

### 4.6. Perbandingan Hasil Pengelompokan dari tahun ke Tahun

Hasil pengelompokan menunjukkan bahwa Indeks validitas *Silhouette* merupakan indeks validitas yang lebih baik dalam penentuan jumlah optimum *cluster* dibandingkan indeks validitas *Gap statistic*. Maka dilakukan perbandingan hasil pengelompokan terhadap hasil pengelompokan dengan menggunakan Indeks validitas *Silhouette* dari Tahun 2005 ke Tahun 2008, dengan jumlah kelompok yang telah ditentukan adalah 3.

Hasil pengelompokan metode *K-Means* dengan menggunakan indeks validitas *Silhouette* dengan *k*=3 pada data Penyusunan Kinerja Makro ekonomi dan Sosial Jawa Timur tentang Pembangunan Manusia dari Tahun 2005 sampai Tahun 2008 secara lengkap dapat dilihat pada Lampiran 6. Dari hasil pengelompokan tersebut diketahui bahwa masing-masing kelompok memiliki ciri-ciri pengelompokan yang membedakan antar kelompok.

Ciri-ciri kelompok 1yaitu angota kelompok yang menjadi anggota kelompok 1 adalah wilayah yang memiliki nilai indeks kesehatan (IHH), indeks pendidikan (IP), daya beli (PPP) dan IPM sedang. Sedangkan pada kelompok 2 diperoleh ciri-ciri setiap wilayah yang menjadi anggota kelompok 2 memiliki nilai indeks kesehatan (IHH), indeks pendidikan (IP), daya beli (PPP) dan IPM yang tinggi. Ciri-ciri dari kelompok 3 adalah anggota kelompok yang diperoleh adalah wilayah yang memiliki nilai indeks kesehatan (IHH), indeks pendidikan (IP), daya beli (PPP) dan IPM rendah. Dalam artian wilayah yang masuk pada kelompok 1 merupakan wilayah yang memiliki nilai indeks kesehatan (IHH), indeks pendidikan (IP), daya beli (PPP) dan IPM lebih rendah daripada kelompok 2 tetapi lebih tinggi jika dibandingkan kelompok 3.

Dari hasil pengelompokan, dapat dikatakan bahwa dalam kelompok yang dihasilkan wilayah yang ada dalam kelompok yang sama memiliki nilai indeks kesehatan (IHH), indeks pendidikan (IP), dan daya beli (PPP) rendah maka Indeks Pembangunan Manusia (PPP) masyarakat di wilayah tersebut pun juga rendah. Sebaliknya, jika suatu wilayah memiliki memiliki nilai indeks kesehatan (IHH), indeks pendidikan (IP), dan daya beli (PPP) tinggi maka Indeks Pembangunan Manusia (PPP) masyarakat di wilayah tersebut pun juga tinggi.

Berdasarkan hasil pengelompokan metode K-Means dengan menggunakan indeks validitas *Silhouette* dari tahun ke tahun

menunjukkan hasil *cluster* yang berbeda-beda. Perbandingan dilakukan untuk mengetahui pergeseran hasil pengelompokan dari tahun ke tahun. Hal ini perlu dilakukan untuk mengetahui perubahan kelompok data Penyusunan Kinerja Makro ekonomi dan Sosial Jawa Timur tentang Pembangunan Manusia dari Tahun 2005 ke Tahun 2008.

Kabupaten/Kota yang masuk sebagai anggota *cluster* 1, *cluster* 2 dan *cluster* 3 pada *K-means cluster* dengan menggunakan indeks validitas *Silhouette* dengan jumlah cluster yang telah ditentukan adalah 3 untuk Data 1, Data 2, Data 3 dan Data 4 disajikan dalam tabel 4.8.

Tabel 4.8. Hasil Pengelompokan dengan Indeks Validitas *Silhouette* untuk *k*=3

Volomenak		Anggota F	Kelompok	1
Kelompok	2005	2006	2007	2008
	Pacitan	Pacitan	Pacitan	Pacitan
	Ponorogo	Ponorogo	Ponorogo	Ponorogo
	Trenggalek	Trenggalek	Trenggalek	Kediri
	Blitar	Tulungagung	Kediri	Malang
	Kediri	Kediri	Malang	Lumajang
	Malang	Malang	Banyuwangi	Banyuwangi
	Banyuwangi	Lumajang	Jombang	Nganjuk
1	Mojokerto	Banyuwangi	Nganjuk	Madiun
	Jombang	Mojokerto	Madiun	Ngawi
	Nganjuk	Jombang	Magetan	Bojonegoro
32.1	Madiun	Nganjuk	Ngawi	Tuban
	Magetan	Madiun	Bojonegoro	Lamongan
450 1	Ngawi	Magetan	Tuban	K. Batu
AHT.	Bojonegoro	Ngawi	Lamongan	
N. LA-FT	Tuban	Bojonegoro		

Tabel 4.8. (Lanjutan)

Kel.	Anggota Kelompok					
Kei.	2005	2006	2007	2008		
	Lamongan	Tuban		TO A		
1	K. Probolinggo	Lamongan				
HT.	K. Pasuruan	K. Pasuruan				
	Tulungagung	Blitar	Tulungagung	Trenggalek		
	Sidoarjo	Sidoarjo	Blitar	Tulungagung		
	Gresik	Gresik	Sidoarjo	Blitar		
	K. Kediri	K. Kediri	Mojokerto	Sidoarjo		
	K.Blitar	K. Blitar	Gresik	Mojokerto		
	K. Malang	K. Malang	K. Kediri	Jombang		
	K. Mojokerto	K. Probolinggo	K. Blitar	Magetan		
2	K. Madiun	K. Mojokerto	K. Malang	Gresik		
2	K. Surabaya	K. Madiun	K. Probolinggo	K. Kediri		
	K. Batu	K. Surabaya	K. Pasuruan	K. Blitar		
	7	K. Batu	K. Mojokerto	K. Malang		
			K. Madiun	K. Probolinggo		
			K. Surabaya	K. Pasuruan		
			K. Batu	K. Mojokerto		
		PAST CO		K. Madiun		
				K. Surabaya		

Tabel 4.8. (Lanjutan)

Kel.	LOAW	Anggota	Kelompok	
Kel.	2005	2006	2007	2008
	Lumajang	Jember	Lumajang	Jember
	Jember	Bondowoso	Jember	Bondowoso
	Bondowoso	Situbondo	Bondowoso	Situbondo
	Situbondo	Probolinggo	Situbondo	Probolinnggo
3	Probolinggo	Pasuruan	Probolinggo	Pasuruan
3	Pasuruan	Bangkalan	Pasuruan	Bangkalan
	Bangkalan	Sampang	Bangkalan	Sampang
	Sampang	Pamekasan	Sampang	Pamekasan
	Pamekasan	Sumenep	Pamekasan	Sumenep
	Sumenep	人又多	Sumenep	

Tabel 4.8 menunjukkan bahwa Kabupaten/Kota yang masuk sebagai anggota dalam *cluster* 1 dari tahun ke tahun terjadi pergeseran. Tahun 2005 anggota *cluster* 1 sebanyak 18 Kabupaten/Kota demikian juga pada tahun 2006 anggota *cluster* 1 sebanyak 18 Kabupaten/Kota. Kabupaten Blitar dan Kota Probolinggo yang pada Tahun 2005 masuk pada *cluster* 1, pada Tahun 2006 masuk menjadi anggota *cluster* 2. Sedangkan Kabupaten Tulungagung dan Kab. Lumajang yang pada Tahun 2006 masuk pada *cluster* 1, pada Tahun 2005 secara berturut-turut masuk menjadi anggota *cluster* 1 dan *cluster* 3.

Tahun 2007 anggota *cluster* 1 sebanyak 14 Kabupaten. Pada tahun 2007, Kab.Tulungagung, Kab. Mojokerto dan Kota Pasuruan yang pada tahun sebelumnya berada pada *cluster* 1 berubah menjadi anggota *cluster* 2. Sedangkan Kab. Lumajang yang pada Tahun 2006 masuk menjadi anggota *cluster* 1 di tahun 2007 menjadi anggota *cluster* 3.

Tahun 2008 anggota cluster 1 sebanyak 13 Kabupaten/Kota. Kab.Trenggalek, Kab. Jombang, dan Kab. Magetan yang pada Tahun 2007 masuk pada *cluster* 1, di tahun 2008 mengalami pergeseran menjadi anggota kelompok 2. Pada Tahun 2008 Kab. Lumajang dan Kota Batu merupakan anggota kelompok 1,

sedangkan pada Tahun 2007 Kab. Lumajang merupakan anggota kelompok 3, dan Kota Batu merupakan anggota kelompok 2.

Hasil kelompok 2 yang ditunjukkan Tabel 4.5 terlihat anggota kelompok 2 dari tahun ke tahun terlihat mengalami peningkatan jumlah anggota kelompok. Pada Tahun 2005, jumlah anggota cluster 2 sebanyak 10 Kabupaten/Kota. Pada Tahun 2006, jumlah anggota cluster 2 menjadi sebanyak 11 Kabupaten/Kota, karena Kabupaten Blitar dan Kota Probolinggo yang pada Tahun 2005 masuk pada *cluster* 1, pada Tahun 2006 masuk menjadi anggota *cluster* 2. Sedangkan Kabupaten Tulungagung yang pada Tahun 2005 masuk pada *cluster* 2, pada Tahun 2006 bergeser menjadi anggota *cluster* 1.

Kab. Tulungagung, Kab. Mojokerto dan Kota Pasuruan pada yang pada tahun sebelumnya berada pada cluster 1, di Tahun 2007 masuk pada cluster 2. Sehingga jumlah anggota cluster 2 meningkat menjadi sebanyak 14 Kabupaten/kota. Tahun 2008 jumlah anggota cluster 2 meningkat menjadi sebanyak 16 Kabupaten/Kota. Karena Kab. Trenggalek, Kab Jombang dan Kab. Magetan masuk pada cluster 2.

Dari Tabel 4.8 diketahui bahwa anggota kelompok 3 pada Tahun 2005 sebanyak 10 Kabupaten. Tahun 2006 anggota kelompok 3 menjadi 9 Kabupaten karena Kab. Lumajang keluar dari anggota kelompok 3. Tahun 2007 anggota kelompok 3 kembali memiliki 10 anggota kelompok karena Kab. Lumajang kembali masuk pada cluster 3. Untuk tahun 2008 anggota kelompok 3 sebanyak 9 Kabupaten karena Kab. Lumajang kembali keluar dari anggota kelompok 3. Anggota kelompok 3 pada Tahun 2005 sama dengan anggota kelompok 3 pada Tahun 2006 sama dengan anggota kelompok 3 pada Tahun 2006 sama dengan anggota kelompok 3 pada Tahun 2008.

Berdasarkan hasil pengelompokan metode K-Means dengan menggunakan indeks validitas Silhouette dengan k=3, diperoleh informasi bahwa terdapat beberapa Kabupaten/Kota yang dari tahun ke tahun mengalami pergeseran keanggotaan kelompok. Misalnya, Kab. Blitar pada Tahun 2005 merupakan anggota cluster 1. Tetapi, pada Tahun 2006, Tahun 2007 dan Tahun 2008 Kab. Blitar menjadi anggota cluster 2. Demikian juga dengan Kota Probolinggo, pada Tahun 2005 merupakan anggota cluster 1 tetapi pada Tahun 2006, Tahun 2007 dan Tahun 2008 menjadi

anggota cluster 2. Sedangkan Kab. Tulungagung yang pada Tahun 2005 merupakan anggota cluster 2, di Tahun 2006 menjadi anggota cluster 1 dan di tahun-tahun selanjutnya kembali menjadi anggota cluster 2. Kota Batu pada Tahun 2005, Tahun 2006 dan Tahun 2007 berada pada cluster 2, tetapi pada tahun 2008 masuk menjadi anggota cluster 1.

Kabupaten Lumajang adalah Kabupaten yang dari tahun ke tahun mengalami perubahan cluster yang mencolok. Pada Tahun 2005 Kabupaten Lumajang menjadi anggota cluster 3, sedangkan di Tahun 2006 menjadi anggota cluster 1. Sebaliknya, pada tahun 2007 Kabupaten Lumajang kembali menjadi anggota cluster 3, sedangkan di Tahun 2008 kembali menjadi anggota cluster 1.

Berbeda dengan Kota Pasuruan. Kota ini pada Tahun 2005 dan Tahun 2006 berada pada cluster 1, sedangkan pada Tahun 2007 dan Tahun 2008 menjadi anggota cluster 2. Dan seterusnya, terdapat beberapa kabupaten lain yang dari tahun ke tahun cenderung mengalami pergeseran keanggotaan pengelompokan.

Tetapi, terdapat beberapa Kabupaten/kota yang dari tahun ke tahun tidak mengalami pergeseran keanggotaan kelompok. Seperti Kab. Pacitan, Kab. Ponorogo, Kab. Kediri, Kab. Malang dan Kab. Banyuwangi dari tahun ke tahun berada pada cluster yang tetap, yaitu berada pada cluster 1.

Berdasarkan hasil pengelompokan diketahui pergeseran keanggotaan kelompok dari Tahun 2005-2008 serta diketahui wilayah mana yang masuk pada kategori memiliki pembangunan manusia rendah, sedang maupun tinggi. Pemerintah harus memfokuskan perhatian pada wilayah yang masuk pada wilayah dengan kategori pembangunan manusia yang rendah yaitu daerahdaerah yang memiliki nilai indeks kesehatan (IHH), indeks pendidikan (IP), daya beli (PPP) dan IPM rendah.

Pemerintah harus mengarahkan perhatian pada wilayah yang masuk pada kelompok 3. Karena, wilayah yang masuk pada kelompok 3 adalah wilayah yang masih memiliki indeks kesehatan, indeks pendidikan, daya beli dan IPM yang paling rendah dibandingkan kelompok yang lain. Berdasarkan hasil yang diperoleh dari Analisis Komponen Utama, terdapat dua variabel yang memiliki bobot tertinggi yaitu Indeks Harapan Hidup (IHH) dan Indeks Pendidikan (IP). Maka, perhatian pemerintah seharusnya lebih difokuskan pada kesehatan dan pendidikan

terhadap wilayah yang masuk pada kelompok 3 yaitu dengan pembangunan sarana kesehatan dan sarana pendidikan yang memadai. Selain itu masyarakat yang berada di daerah tersebut sangat membutuhkan adanya pembinaan terhadap pola pikir mereka tentang pentingnya pemanfaatan sarana kesehatan dan sarana pendidikan secara optimal.





### BAB V KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Berdasarkan hasil penelitian ini dapat diambil kesimpulan bahwa:

- 1. Hasil pengelompokan menggunakan indeks validitas *Gap Statistic* pada metode *K-means* menghasilkan jumlah kelompok optimum yang sama pada data Tahun 2005, Tahun 2006, Tahun 2007 dan Tahun 2008, yaitu didapatkan jumlah kelompok optimum sama dengan 3.
- 2. Hasil pengelompokan menggunakan indeks validitas Silhouette pada metode *K-means* menghasilkan jumlah kelompok optimum yang berbeda-beda dari keempat data penelitian. Data Tahun 2005 dan Data Tahun 2006 menghasilkan jumlah kelompok optimum sebanyak 3 *cluster*, sedangkan pada Data Tahun 2007 dan Data Tahun 2008 diperoleh jumlah kelompok optimum adalah 5 *cluster*.
- 3. Berdasarkan hasil nilai CTM terkecil, dari keseluruhan hasil pengelompokan menunjukkan bahwa Indeks validitas Silhouette menghasilkan nilai CTM terkecil. Jadi, hasil pengelompokan menunjukkan bahwa indeks validitas Silhouette merupakan indeks validitas yang lebih baik dalam penentuan jumlah *cluster* optimum dibandingkan indeks validitas *Gap Statistic* pada metode *K-means Cluster*.
- 4. Perbandingan hasil pengelompokan pada indeks validitas *Silhouette* dengan *k*=3 dari Tahun 2005 ke Tahun 2008 menunjukkan pergeseran hasil pengelompokan. Pemerintah harus mengarahkan perhatian pada wilayah yang masuk pada kelompok 3. Karena, wilayah yang masuk pada kelompok 3 adalah wilayah yang memiliki indeks kesehatan, indeks pendidikan, daya beli dan IPM yang paling rendah dibandingkan kelompok yang lain.

## 5.2 Saran

Untuk penelitian selanjutnya dapat mengkaji tentang:

- 1. Perbandingan indeks validitas *Gap Statistic* dan Silhouette dengan menggunakan metode pengelompokan non hirarki yang lain, seperti *Fuzzy C-Means Cluster* ataupun dengan metode lainnya.
- 2. Penentuan jumlah kelompok optimum dengan menggunakan metode *Compatible Cluster Merging*.



#### DAFTAR PUSTAKA

- Afifi, A. A dan V. Clark. 1990. Computer Aided Multivariate Analysis. Chapman & Hall. New York.
- Arima, C., K. Hakamada, M. Okamoto, dan T. Hanai. 2008. Validity Index for Fuzzy K-Means Clustering Using the Gap statistic Method. Laboratory for Biometrics Graduates School Of System Life Sciences Kyusu University. Japan.
- Bolshakova, N. dan F. Azuaje. 2008. Cluster Validation Techniques for Genome Expression Data. <a href="http://www.cs.tcd.ie/publications/techreports.02/TCD-CS-2002-33.pdf">http://www.cs.tcd.ie/publications/techreports.02/TCD-CS-2002-33.pdf</a>. Tanggal Akses 19 Desember 2010.
- BPS. 2008 . Analisa Penyusunan Kinerja Makro ekonomi dan Sosial Jawa Timur Tahun 2008. BPS Jawa Timur.
- Dodge, Y. 2008. *The Concise Encyclopedia of Statistics*. Springer Science +Business Media, LLC. Berlin.
- Dillon, W. R dan M Goldstein. 1984. *Multivariate Analysis: Methods & Applications*. John Wiley and Sons. New York.
- Dubes dan Jain, A. K. 1988. *Algorithm for Clustering Data*. Prentice Hall. New Jersey.
- Epps, J dan E. Mbikairajah. 2008. Visualisation of Reduced Dimension Microarry Data Using Gaussian Mixture Model. <a href="http://crpit.com/confpapers/CRPITV45Epps.pdf">http://crpit.com/confpapers/CRPITV45Epps.pdf</a>. Tanggal Akses 19 Desember 2010.
- Handayani, P. 2002. Pengelompokan Propinsi di Indonesia Berdasarkan Beberapa Indikator Kesehatan. <a href="http://www.repository.ipb.ac.id/handle/123456789/21941?show=full.pdf">http://www.repository.ipb.ac.id/handle/123456789/21941?show=full.pdf</a>. Tanggal Akses 19 Desember 2010.

- Hair, J. F., R. E. Anderson, R. L. Thatham dan W. C. Black. 1998. Multivariate Data Analysis Fifth Edition. Prentice-Hall, Inc. New Jersey.
- Johnson, R. A. dan D. W. Wichern. 2002. Applied Multivariate Statistical Analysis Fifth Edition. Prentice- Hall, Inc. New Jersey.
- Kariyam dan Subanar. 2002. Perbandingan Beberapa Indeks Kriteria Penentuan Jumlah Optimal Kelompok. Program Studi Matematika. Program Pascasarjana Universitas Gadjah Mada Yogyakarta.
- Kenesei, T., B. Balasko, dan J. Abonyi. 2004. *A Matlab Toolbox and its Web based Variant for Fuzzy Cluster Analysis*. University of Pannonia, Department of Process Engineering. Veszprem, Hungary.
- Kirana, D. A. 2008. Pengelompokan Terbaik Berdasarkan Algoritma K-Means dan Fuzzy C-Means pada Analisis Kelompok Non-Hirarki. Program Studi Statistika Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Brawijaya. Tidak dipublikasikan.
- Laboratorium Data Mining Jurusan Teknik Industri Fakultas Teknologi Industri Universitas Islam Indonesia. 2009. Clustering.

  <a href="http://www.adadata.id.or.id/dl\_jump.php?id=49&Login=31c0">http://www.adadata.id.or.id/dl\_jump.php?id=49&Login=31c0</a>
  <a href="http://www.adadata.id.or.id/dl\_jump.php?id=49&Login=31c0">97c3fb5534f53825c99e05a4cbe5</a>.

  Tanggal Akses 19
  Desember 2010.
- Lattin, J. M., J. D. Carroll, dan P. E. Green. 2006. *Analyzing Multivariate Data*. Duxbury Press. South Western.
- Marriott, F. H. C. 1971. *Practical Problems in a Method of Cluster Analysis*. Biometrics 27, 501-514.
- Pramono, M. S. 2005. Estimasi jumlah Kelompok pada Analisis Kelompok Metode Single Linkage dan K-Means Melalui

- Beberapa Metode Penentuan Jumlah Kelompok. Jurusan Statistika Program Pascasarjana Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Teknologi Sepuluh Nopember Surabaya. Tidak dipublikasikan.
- Rousseeuw, P. J. 1987. Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. Elevier Science Publishers B. V. Journal of Computational and Applied Mathematics 20, 53-65.
- Sharma, S. 1996. *Applied Multivariate Techniques*. John Wiley and Sons, Inc. New York.
- Sukmaraga, P. 2011. Analisis Pengaruh Indeks Pembangunan Manusia, PDRB Per Kapita, Dan Jumlah Pengangguran Terhadap Jumlah Penduduk Miskin Di Provinsi Jawa Tengah. <a href="http://www.undip.id.or.id/dl\_jump.php?id=49&Login=44d088">http://www.undip.id.or.id/dl\_jump.php?id=49&Login=44d088</a> <a href="https://www.undip.id.or.id/dl\_jump.php?id=49&Login=44d088">http://www.undip.id.or.id/dl\_jump.php?id=49&Login=44d088</a> <a href="https://www.undip.id.or.id/dl\_jump.php?id=49&Login=44d088">http://www.undip.id.or.id/dl\_jump.php?id=49&Login=44d088</a> <a href="https://www.undip.id.or.id/dl\_jump.php?id=49&Login=44d088">http://www.undip.id.or.id/dl\_jump.php?id=49&Login=44d088</a> <a href="https://www.undip.id.or.id/dl\_jump.php?id=49&Login=44d088">http://www.undip.id.or.id/dl\_jump.php?id=49&Login=44d088</a> <a href="https://www.undip.id.or.id/dl\_jump.php?id=49&Login=44d088">https://www.undip.id.or.id/dl\_jump.php?id=49&Login=44d088</a> <a href="https://www.undip.id/dl\_jump.php?id=49&Login=44d088">https:
- Tibshirani, W., G. Walther dan T. Hastie. 2001. Estimating The Number of Clusters in A Dataset via The Gap Statistics.

  Technical Report 208 Departement of Statistics Stanford University. Stanford.
- Walpole, R. E. 1992. *Pengantar Statistika, Edisi ke-3*. Terjemahan Sumantri, B. PT. GramediaPustaka Utama. Jakarta.
- Yeung, K. Y. dan W. L. Ruzzo. 2001. Details of the Adjusted Rand index and Clustering algorithms Supplement to the paper "An empirical study on Principal Component Analysis for clustering gene expression data" (to appear in Bioinformatics). <a href="http://www.mendeley.com/research/details-adjusted-rand-index-clustering-algorithms-supplement-paper-empirical-study-principal-component-analysis-clustering-gene-expression-data-appear-bioinformatics/">http://www.mendeley.com/research/details-adjusted-rand-index-clustering-algorithms-supplement-paper-empirical-study-principal-component-analysis-clustering-gene-expression-data-appear-bioinformatics/">http://www.mendeley.com/research/details-adjusted-rand-index-clustering-algorithms-supplement-paper-empirical-study-principal-component-analysis-clustering-gene-expression-data-appear-bioinformatics/">http://www.mendeley.com/research/details-adjusted-rand-index-clustering-algorithms-supplement-paper-empirical-study-principal-component-analysis-clustering-gene-expression-data-appear-bioinformatics/</a>. Tanggal Akses 19 Desember 2010.



## LAMPIRAN

## Lampiran 1. Data Penelitian

Data 1. Data Penyusunan Kinerja Makro ekonomi dan Sosial Jawa Timur tentang Pembangunan Manusia Tahun 2005

No	Kabupaten/Kota	IHH	IP	PPP	IPM
1	Pacitan	75.5	71.35	54.21	67.02
2	Ponorogo	72	67.04	56.4	65.15
3	Trenggalek	75.67	75.44	54.97	68.69
4	Tulungagung	75.67	75.1	58.01	69.59
5	Blitar	74.83	73.47	57.07	68.46
6	Kediri	72.5	75.08	56.11	67.9
7	Malang	70.33	70.68	56.76	65.92
8	Lumajang	68.17	64.03	54.55	62.25
9	Jember /	61.17	65.95	53.02	60.05
10	Banyuwangi	67.67	70.2	55.7	64.52
11	Bondowoso	60.83	60.82	53.16	58.27
12	Situbondo	61.33	61.79	54.28	59.14
13	Probolinggo	58.33	61.46	54.23	58.01
14	Pasuruan	62.17	72.24	54.86	63.09
15	Sidoarjo	73.5	85.71	57.78	72.33
16	Mojokerto	73.17	77.49	55.84	68.83
17	Jombang	73.33	77.13	56.63	69.03
18	Nganjuk	/ 71	73.52	56.57	67.03
19	Madiun	71.5	72.16	55.99	66.55
20	Magetan	74.83	75.25	56.16	68.75
21	Ngawi	72	67	54.91	64.64
22	Bojonegoro	69.17	69.13	53.97	64.09
23	Tuban	69.17	68.78	55.13	64.36
24	Lamongan	70.67	70.92	56.61	66.06

# Lampiran 1. (Lanjutan)

No	Kabupaten/Kota	IHH	IP	PPP	IPM
25	Gresik	74.67	79.24	59.1	71
26	Bangkalan	61.67	64.72	53.7	60.03
27	Sampang	58	50.93	52.56	53.83
28	Pamekasan	62	65.06	53.07	60.05
29	Sumenep	64	64.88	54.86	61.24
30	Kota Kediri	73.83	83.66	57.21	71.57
31	Kota Blitar	76.67	84.76	58.13	73.19
32	Kota Malang	72.67	85.12	59.48	72.42
33	Kota Probolinggo	72.67	75.86	57.11	68.55
34	Kota Pasuruan	68.33	82.47	56.07	68.96
35	Kota Mojokerto	76	84.74	58.93	73.22
36	Kota Madiun	75	85.75	60.01	73.59
37	Kota Surabaya	74.17	86.68	61.8	74.21
38	Kota Batu	71.83	79.25	57.45	69.51

## Keterangan:

IHH = Indeks Harapan Hidup

IP = Indeks Pendidikan

PPP = Purchasing Power Parity

IPM = Indeks Pembangunan Manusia

## Lampiran 1. (Lanjutan)

Data 2. Data Penyusunan Kinerja Makro ekonomi dan Sosial Jawa Timur tentang Pembangunan Manusia Tahun 2006

No	Kabupaten/Kota	IHH	IP	PPP	IPM
1	Pacitan	75.75	73.2	53.7	67.55
2	Ponorogo	72.58	67.13	56.56	65.43
3	Trenggalek	76.08	76.28	54.4	68.92
4	Tulungagung	75.75	74.94	56.13	68.94
5	Blitar	75.58	75.9	57	69.49
6	Kediri	72.58	75.66	54.76	67.67
7	Malang		73.52	55.23	66.41
8	Lumajang	68.33	69.39	53.91	63.88
9	Jember	61.67	68.44	52.16	60.75
10	Banyuwangi	68.25	71.41	55.3	64.99
11	Bondowoso	61.48	58.54	52.7	57.57
12	Situbondo	61.42	62.4	53.68	59.17
13	Probolinggo	58.67	61.71	53.96	58.11
14	Pasuruan	62.25	72.01	54.69	62.98
15	Sidoarjo	73.75	85.69	56.92	72.12
16	Mojokerto	73.67	75.91	56.01	68.53
17	Jombang	73.83	75.45	55.44	68.24
18	Nganjuk	71.08	73.76	54.49	66.44
19	Madiun	71.83	71.04	54.92	65.93
20	Magetan	74.83	74.86	55.25	68.31
21	Ngawi	72.58	63.7	54.5	63.59
22	Bojonegoro	69.33	70	53.69	64.34
23	Tuban	69.5	68.8	53.58	63.96
24	Lamongan	70.83	71.42	55.7	65.99

# Lampiran 1. (Lanjutan)

No	Kabupaten/Kota	ІНН	IP	PPP	IPM
25	Gresik	75.17	81.78	57.23	71.39
26	Bangkalan	62	65.77	52.61	60.13
27	Sampang	58.5	50.07	52.52	53.7
28	Pamekasan	62.42	64.41	53.2	60.01
29	Sumenep	64.42	61.91	55.15	60.49
30	Kota Kediri	74	82.85	56.3	71.05
31	Kota Blitar	77.08	84.73	58.37	73.4
32	Kota Malang	73.33	87.76	61.08	74.06
33	Kota Probolinggo	73.08	74.77	57	68.28
34	Kota Pasuruan	68.83	81.81	55.43	68.69
35	Kota Mojokerto	76.08	85.5	57.98	73.19
36	Kota Madiun	75.08	84.2	59.39	72.89
37	Kota Surabaya	74.42	85.43	58.69	72.84
38	Kota Batu	72.5	79.86	56.41	69.59

## Keterangan:

IHH = Indeks Harapan Hidup

IP = Indeks Pendidikan

PPP = Purchasing Power Parity

IPM = Indeks Pembangunan Manusia

Data 3. Data Penyusunan Kinerja Makro ekonomi dan Sosial Jawa Timur tentang Pembangunan Manusia Tahun 2007

No	Kabupaten/Kota	IHH	IP	PPP	IPM
1	Pacitan	76.31	76.2	53.81	68.77
2	Ponorogo	73.29	71.43	55.66	66.79
3	Trenggalek	76.5	75.44	55.01	68.98
4	Tulungagung	75.84	79.55	55.31	70.23
5	Blitar	76.02	75.25	56.1	69.13
6	Kediri	72.58	78.7	54.74	68.67
7	Malang	70.46	74.58	55.16	66.73
8	Lumajang	68.5	67.98	52.1	62.86
9	Jember	62.42	68.85	52.46	61.24
10	Banyuwangi	68.23	72.22	54.95	65.13
11	Bondowoso	62.47	60.3	51.14	57.97
12	Situbondo	61.3	66.79	53.37	60.48
13	Probolinggo	58.69	62.6	52.48	57.92
14	Pasuruan	62.33	73.11	51.88	62.44
15	Sidoarjo	73.9	86.07	56.18	72.05
16	Mojokerto	73.78	80.37	56.05	70.07
17	Jombang	74.68	78.73	56.57	70
18	Nganjuk	70.99	75.32	55.34	67.22
19	Madiun	71.66	72.49	55.03	66.39
20	Magetan	74.88	74.77	55.84	68.5
21	Ngawi	72.78	71.13	54.16	66.02
22	Bojonegoro	69.49	71.04	54.92	65.15
23	Tuban	69.83	69.99	53.6	64.47
24	Lamongan	70.9	72.98	54.97	66.28
25	Gresik	75.59	79.86	54.89	70.11
26	Bangkalan	62.33	63.65	52.71	59.56
27	Sampang	59.45	49.91	52.79	54.05

No	Kabupaten/Kota	IHH	IP	PPP	IPM
28	Pamekasan	62.83	65.27	53.23	60.44
29	Sumenep	65.49	64.53	54.61	61.54
30	Kota Kediri	74.13	85.99	57.34	72.49
31	Kota Blitar	77.53	85.54	56.77	73.28
32	Kota Malang	74.23	87.21	58.34	73.26
33	Kota Probolinggo	73.5	80.45	57.17	70.38
34	Kota Pasuruan	68.96	84.07	56.15	69.73
35	Kota Mojokerto	76.26	86.27	58.38	73.64
36	Kota Madiun	75.35	88.46	59.28	74.36
37	Kota Surabaya	74.5	87.96	58.26	73.57
38	Kota Batu	73.33	83.73	55.46	70.84

# Keterangan:

= Indeks Harapan Hidup IHH

ΙP = Indeks Pendidikan

PPP

= Purchasing Power Parity = Indeks Pembangunan Manusia IPM

Data 4. Data Penyusunan Kinerja Makro ekonomi dan Sosial Jawa Timur tentang Pembangunan Manusia Tahun 2008

No	Kabupaten/Kota	IHH	IP	PPP	IPM
1	Pacitan	77.35	76.26	54.96	69.52
2	Ponorogo	74.33	71.47	58.71	68.17
3	Trenggalek	77.6	77.65	57.81	71.02
4	Tulungagung	76.79	79.21	60.21	72.07
5	Blitar	77.38	76.83	58.38	70.87
6	Kediri	73.24	77.92	56.67	69.28
7	Malang	70.92	74.51	56.92	67.45
8	Lumajang	69.32	71.11	55.02	65.15
9	Jember	63.32	68.87	54.8	62.33
10	Banyuwangi	69.21	73.05	58.39	66.88
11	Bondowoso	63.63	61.58	52.15	59.12
12	Situbondo	61.59	64.52	54.49	60.2
13	Probolinnggo	59.09	63.77	53.82	58.89
14	Pasuruan	62.76	73.55	55.89	64.07
15	Sidoarjo	74.58	86.27	59.06	73.3
16	Mojokerto	74.49	79.68	58.53	70.9
17	Jombang	75.74	78.84	59.49	71.36
18	Nganjuk	71.27	76.72	58.12	68.7
19	Madiun	72.55	74.59	59.58	68.91
20	Magetan	75.1	77.04	59.06	70.4
21	Ngawi	73.7	70.79	56.01	66.83
22	Bojonegoro	70.19	70.48	57.95	66.21
23	Tuban	70.47	70.55	57.89	66.3
24	Lamongan	71.4	73.75	58.8	67.98
25	Gresik	76.37	81.31	58.92	72.2
26	Bangkalan	62.79	65.75	57.26	61.94
27	Sampang	60.74	51.38	55.17	55.77

No	Kabupaten/Kota	IHH	IP	PPP	IPM
28	Pamekasan	63.26	66.58	53.91	61.25
29	Sumenep	66.51	64.31	55.63	62.15
30	Kota Kediri	74.71	86.8	58.66	73.39
31	Kota Blitar	78.5	86.59	59.82	74.97
32	Kota Malang	75.36	88.37	55.64	73.13
33	Kota Probolinggo	74.23	79.59	58.52	70.87
34	Kota Pasuruan	69.91	83.59	60.76	71.42
35	Kota Mojokerto	77.22	86.33	61.18	74.91
36	Kota Madiun	75.77	87.85	61.15	74.93
37	Kota Surabaya	75.09	87.96	65.24	76.1
38	Kota Batu	73.94	84	53.5	70.48

# Keterangan:

IHH = Indeks Harapan Hidup

IP = Indeks Pendidikan

PPP = Purchasing Power Parity

IPM = Indeks Pembangunan Manusia

## Lampiran 2. Uji Koefisien Korelasi

#### Data 1

## Correlations: IHH, IP, PPP, IPM

IHH IP PPP

1P 0.779
0.000

PPP 0.730 0.852
0.000 0.000

IPM 0.907 0.967 0.886
0.000 0.000 0.000

Cell Contents: Pearson correlation P-Value

#### Data 2

## Correlations: IHH, IP, PPP, IPM

IHH IP PPP

IP 0.775
0.000

PPP 0.680 0.802
0.000 0.000

IPM 0.904 0.966 0.844
0.000 0.000 0.000

Cell Contents: Pearson correlation P-Value

#### Data 3

# Correlations: IHH, IP, PPP, IPM

IHH IP PPP

1P 0.797
0.000

PPP 0.757 0.835
0.000 0.000

IPM 0.915 0.969 0.881
0.000 0.000 0.000

Cell Contents: Pearson correlation P-Value

AWILLA

#### Data 4

## Correlations: IHH, IP, PPP, IPM

IHH IP PPP

IP 0.799
0.000

PPP 0.606 0.637
0.000 0.000

IPM 0.916 0.961 0.749
0.000 0.000 0.000

Cell Contents: Pearson correlation

P-Value

## Lampiran 3. Hasil Analisis Komponen Utama

## Data 1

# Principal Component Analysis: IHH, IP, PPP, IPM

Eigenanalysis of the Correlation Matrix

Eigenvalue 3.5660 0.2847 0.1493 0.0000 Proportion 0.892 0.071 0.037 0.000 Cumulative 0.892 0.963 1.000 1.000

Variable PC1 PC2 PC3 PC4 IHH -0.479 0.778 0.272 0.303 -0.506 -0.253 -0.682 ΙP 0.463 PPP -0.486 -0.572 0.651 0.115 IPM -0.528 0.064 -0.192 -0.825

#### Skor Komponen Utama Data 1

No	Skor PC1	Skor PC2	Skor PC3	Skor PC4
1	-0.01441	1.36673	-0.205735	0.0002622
2	0.23678	0.38323	0.721856	-0.0007533
3	-0.62671	1.08219	-0.356974	0.0007956
4	-1.40576	0.276	0.57855	0.0002687
5	-0.89825	0.44828	0.421135	-0.0007087
6	-0.51157	0.32628	-0.099956	-0.0004818
7	-0.00162	-0.04933	0.427084	0.0003991
8	1.48346	0.40133	0.314914	0.0000561
9	2.56088	-0.25301	-0.57473	-0.0001237
10	0.65075	-0.13852	0.060211	0.0005868
11	3.0533	-0.20768	-0.064387	0.00026
12	2.60052	-0.46033	0.195529	-0.0010054
13	3.01078	-0.87349	0.102574	-0.0004336

No	Skor PC1	Skor PC2	Skor PC3	Skor PC4
14	1.35	-0.76341	-0.581	0.0002812
15	-2.08997	-0.2504	-0.563662	0.0000741
16	-0.74964	0.43362	-0.381532	0.0007248
17	-0.94607	0.25436	-0.107257	0.0000299
18	-0.303	0.02561	0.128588	-0.0000431
19	-0.07967	0.28865	0.101884	0.0000135
20	-0.82471	0.64636	-0.016251	-0.0005038
21	0.63827	0.7837	0.282701	-0.0005008
22	1.03135	0.57038	-0.299264	0.0002903
23	0.75504	0.26843	0.078363	0.0000831
24	-0.02545	0.03396	0.372494	0.0009777
25	-1.96883	-0.2679	0.478016	0.0002074
26	2.43589	-0.33114	-0.238956	0.0002712
27	4.49881	-0.20232	0.58154	0.0001062
28	2.53082	-0.12308	-0.446351	-0.0007128
29	1.82814	-0.30845	0.175892	0.0011581
30	-1.78333	0.00313	-0.529016	-0.0004382
31	-2.47922	0.13998	-0.255337	-0.0005861
32	-2.3862	-0.81132	-0.03309	0.0003238
33	-0.87338	0.06275	0.130352	-0.0006314
34	-0.69636	-0.45813	-0.95644	-0.0002345
35	-2.60859	-0.17121	-0.039783	0.0003806
36	-2.87192	-0.63168	0.149912	-0.0008652
37	-3.33607	-1.25599	0.564983	0.0005101
38	-1.18404	-0.23761	-0.116854	-0.0000383

#### Data 2

# Principal Component Analysis: IHH, IP, PPP, IPM

Eigenanalysis of the Correlation Matrix

Eigenvalue 3.4931 0.3262 0.1806 0.0000 Proportion 0.873 0.082 0.045 0.000 Cumulative 0.873 0.955 1.000 1.000

PC3 Variable PC1 PC2 PC4 -0.481 0.691 0.447 0.303 IHH ΙP -0.509 -0.099 -0.714 0.471 PPP -0.475 -0.709 0.510 0.110 -0.533 0.103 -0.176 -0.821 IPM

## Skor Komponen Utama Data 2

No	Skor PC1	Skor PC2	Skor PC3	Skor PC4
1	-0.20685	1.32723	-0.03454	-0.0000831
2	-0.02161	-0.05634	1.01661	-0.0008701
3	-0.73118	1.11371	-0.13469	-0.0000872
4	-1.03551	0.47601	0.38967	0.0000762
5	-1.34295	0.14711	0.49782	0.0006729
6	-0.34045	0.52847	-0.23124	-0.0006482
7	-0.00852	-0.09981	-0.0576	0.001052
8	1.01006	0.28998	-0.13603	-0.0006485
9	2.39818	0.01982	-0.93193	0.0007764
10	0.44852	-0.21229	0.00363	-0.0005804
11	3.21451	-0.14659	0.12727	0.0004789
12	2.58688	-0.51237	-0.00639	-0.0006242
13	2.91493	-0.97046	-0.06333	0.0004699

No	Skor PC1	Skor PC2	Skor PC3	Skor PC4
14	1.29664	-0.79777	-0.61737	0.0003642
15	-2.02668	-0.11282	-0.57963	-0.0001201
16	-0.83911	0.23781	0.12404	0.0000138
17	-0.65949	0.4589	0.04056	-0.00002
18	0.09899	0.43229	-0.21988	0.0004465
19	0.14748	0.39509	0.19503	0.0000013
20	-0.67419	0.65986	0.12004	0.0005421
21	0.86771	0.67416	0.84326	0.000685
22	0.88952	0.49578	-0.17795	-0.0001144
23	1.01266	0.56203	-0.07873	-0.0000951
24	0.02072	-0.00954	0.27858	-0.0010581
25	-1.91404	-0.01431	-0.03397	0.0005597
26	2.4874	-0.08001	-0.54629	-0.0007736
27	4.43427	-0.43907	0.68213	-0.0004997
28	2.40428	-0.22287	-0.24454	-0.000141
29	1.86393	-0.62317	0.60524	0.0006502
30	-1.61832	0.14859	-0.44267	-0.0001058
31	-2.74151	-0.17054	0.09426	-0.0010302
32	-3.30763	-1.62136	0.20376	-0.0003839
33	-0.92811	-0.17856	0.43188	0.0006586
34	-0.64617	-0.22905	-0.91348	-0.0002334
35	-2.58488	-0.17124	-0.14285	-0.000544
36	-2.7229	-0.78686	0.2537	0.0001318
37	-2.56681	-0.6372	-0.07868	0.0011378
38	-1.17975	-0.07422	-0.23568	-0.0000562

#### Data 3

# Principal Component Analysis: IHH, IP, PPP, IPM

Eigenanalysis of the Correlation Matrix

Eigenvalue 3.5812 0.2506 0.1682 0.0000 Proportion 0.895 0.063 0.042 0.000 Cumulative 0.895 0.958 1.000 1.000

Variable PC1 PC2 PC3 PC4 -0.484 0.750 0.341 0.295 IHH ΙP -0.504 -0.196 -0.695 0.474 PPP -0.485 -0.628 0.600 0.105 -0.526 0.076 -0.201 -0.823 IPM

## Skor Komponen Utama Data 3

No	Skor PC1	Skor PC2	Skor PC3	Skor PC4
1	-0.46701	1.19246	-0.19944	0.0003389
2	-0.18114	0.25795	0.63928	0.0005162
3	-0.761	0.85123	0.23466	0.0004651
4	-1.14297	0.59076	-0.08909	0.0004703
5	-0.99472	0.43999	0.5502	-0.0010505
6	-0.5015	0.32334	-0.33941	0.000384
7	0.01949	-0.04	0.0611	0.0004234
8	1.73606	0.76861	-0.32948	-0.0004155
9	2.30306	-0.22512	-0.60273	0.0000964
10	0.57087	-0.24926	0.10741	0.0003904
11	3.45774	0.34994	-0.2001	-0.0005385
12	2.37267	-0.63786	-0.19851	0.0007265
13	3.33182	-0.65319	-0.20278	0.0000888

No	Skor PC1	Skor PC2	Skor PC3	Skor PC4
14	2.08644	-0.12741	-1.17238	-0.0005224
15	-1.75109	-0.07457	-0.53034	-0.000031
16	-1.17535	0.04863	-0.04795	-0.0005604
17	-1.28321	0.03977	0.30142	-0.0010321
18	-0.16564	-0.03462	0.07171	-0.0006373
19	0.10087	0.20814	0.27495	0.0004317
20	-0.73603	0.36888	0.46176	-0.0005455
21	0.33491	0.66745	0.19878	0.0003613
22	0.53241	-0.04012	0.26951	-0.0001352
23	0.96203	0.44594	-0.00652	0.0002804
24	0.16642	0.11069	0.17452	0.0004158
25	-1.02152	0.68335	-0.25412	0.0004242
26	2.72163	-0.22716	-0.05269	0.0001438
27	4.31743	-0.42351	1.09898	-0.0003562
28	2.36345	-0.34948	-0.02419	0.000194
29	1.71233	-0.39722	0.58284	0.000363
30	-2.10156	-0.40915	-0.16907	-0.0004519
31	-2.31684	0.26303	-0.12835	0.0000672
32	-2.50953	-0.73401	0.01864	0.0001848
33	-1.46637	-0.3485	0.26161	-0.0009912
34	-0.95048	-0.73241	-0.59768	-0.0006151
35	-2.68477	-0.44184	0.21714	-0.0003271
36	-3.02886	-0.89534	0.23607	0.0008257
37	-2.58874	-0.68326	-0.06089	0.0007166
38	-1.2613	0.11387	-0.55482	-0.0000988

#### Data 4

# Principal Component Analysis: IHH, IP, PPP, IPM

Eigenanalysis of the Correlation Matrix

Eigenvalue 3.3511 0.4467 0.2022 0.0000 Proportion 0.838 0.112 0.051 0.000 Cumulative 0.838 0.949 1.000 1.000

Variable PC1 PC2 PC3 PC4 -0.499 -0.358 0.730 0.299 IHH ΙP -0.511 -0.272 -0.672 0.462 PPP -0.440 0.883 0.074 0.141 -0.544 -0.131 -0.099 -0.823 IPM

## Skor Komponen Utama Data 4

No	Skor PC1	Skor PC2	Skor PC3	Skor PC4
1	-0.29371	-1.33574	0.635372	0.0009833
2	-0.22278	0.31577	0.745581	0.0005484
3	-1.04063	-0.47076	0.610734	0.0005959
4	-1.578	0.31676	0.428776	0.0007171
5	-1.05183	-0.23369	0.665199	-0.0004666
6	-0.2848	-0.53802	0.015363	-0.0000442
7	0.28216	-0.14745	0.019698	0.0004522
8	1.19648	-0.519	0.066528	0.0003125
9	2.21081	-0.06106	-0.501052	0.0002166
10	0.33684	0.52173	-0.038604	0.0010316
11	3.40926	-0.66247	0.100126	0.0000265
12	2.90729	0.13959	-0.354976	0.0001438
13	3.43096	0.13239	-0.619639	0.0006177

No	Skor PC1	Skor PC2	Skor PC3	Skor PC4
14	1.61144	0.14989	-0.945741	-0.000235
15	-1.73837	-0.18601	-0.474377	0.0012411
16	-0.98995	-0.08755	0.063297	0.0006152
17	-1.26368	0.17124	0.312618	0.0001274
18	-0.21776	0.13319	-0.098074	0.0010739
19	-0.47464	0.60674	0.275618	0.0000589
20	-0.92303	0.14931	0.376059	0.0006237
21	0.47355	-0.50063	0.665385	0.0009452
22	0.54799	0.40866	0.293328	-0.0000823
23	0.51902	0.36569	0.321347	0.0009987
24	-0.08954	0.46804	0.185773	0.0010815
25	-1.46199	-0.16274	0.169198	0.000669
26	2.07201	0.91495	-0.248538	-0.0007787
27	4.12912	0.95768	0.670876	-0.0010309
28	2.61886	-0.25742	-0.333223	0.0001524
29	2.07486	0.16329	0.305537	0.0002504
30	-1.72396	-0.34895	-0.51195	0.0006878
31	-2.41819	-0.23583	0.007122	0.0007831
32	-1.33928	-1.456	-0.630035	-0.0005172
33	-0.95621	-0.07047	0.036336	-0.0139732
34	-1.24221	0.82518	-0.795639	0.0007082
35	-2.5101	0.31692	-0.101746	0.0008325
36	-2.46721	0.35163	-0.413896	-0.0002514
37	-3.2275	1.74558	-0.419178	0.0004799
38	-0.30527	-1.88045	-0.483201	0.0004051

Lampiran 4 Nilai indeks validitas *Gap Statistic, diffu* dan *cluster membership* hasil output *software* R.2.13.0

#### Data 1

#### K=3

\$gap

[1] 0.5686075

\$diffu

[1] 0.08382931

#### K=4

\$gap

[1] 0.6782011

\$diffu

[1] 0.3805573

[1] 1 2 3 2 2 1 2 2 3 4 1 4 4 3 3 3 1 4 4 1 4 1 4 2 2 1 2 3 4 3 1 1 4 3 [35] 1 4 2 1

#### K=5

\$gap

[1] 0.5663874

\$diffu

[1] 0.08772923

[1] 1 2 1 2 2 3 2 2 4 5 3 5 5 4 4 1 3 5 5 3 5 1 5 2 2 1 2 4 5 4 1 3 5 4 [35] 3 5 2 3

# Lampiran 4. (Lanjutan) **Data 2**

#### K=3

\$gap [1] 0.6002252

Śdiffu

[1] 0.1820591

#### K=4

\$gap [1] 0.6436619

\$diffu

[1] 0.2396834

[1] 1 2 1 3 2 1 1 1 4 1 3 1 1 4 4 3 1 1 3 3 2 1 1 3 1 4 2 1 2 4 3 3 3 4 [35] 1 3 1 1

#### K=5

\$gap

[1] 0.543767

Śdiffu

[1] 0.1155949

[1] 1 2 3 4 2 3 1 3 5 1 4 1 1 5 5 4 1 3 4 4 2 3 1 4 1 5 2 3 2 5 4 4 4 5 [35] 3 4 1 3

# Lampiran 4. (Lanjutan) **Data 3**

#### K=3

\$gap

[1] 0.2968127

\$diffu

[1] 0.05085269

#### K=4

\$gap

[1] 0.3469865

\$diffu

[1] 0.1727623

[1] 1 2 3 1 2 1 3 1 4 3 1 1 1 4 4 1 3 3 3 2 3 3 1 3 1 1 2 1 2 1 1 1 1 3 4 [35] 3 3 1 4

#### K=5

\$gap

[1] 0.3543581

\$diffu

[1] 0.1950134

[1] 1 2 3 4 2 1 4 1 5 3 1 1 1 5 5 4 3 4 3 2 3 3 4 3 1 4 2 4 2 1 1 4 3 5 [35] 3 3 4 5

# Lampiran 4. (Lanjutan) **Data 4**

## K=3

\$gap

[1] 0.6095027

\$diffu

[1] 0.1098327

#### K=4

\$gap

[1] 0.5171042

\$diffu

[1] 0.1760587

[1] 1 1 1 2 1 3 3 3 4 3 3 4 4 4 4 3 2 3 2 2 1 2 2 2 2 4 1 4 2 4 3 4 3 4 [35] 3 4 4 4

#### K=5

\$gap

[1] 0.572353

\$diffu

[1] 0.172062

[1] 1 1 1 2 1 3 3 3 4 3 3 4 5 5 4 3 2 3 2 2 1 2 2 2 2 4 1 4 2 4 3 5 3 5 [35] 3 4 4 4

Lampiran 5 Nilai indeks validitas *Silhouette* dan *cluster membership* hasil output *software* R.2.13.0

#### Data 1

K=3

- [1] "max S for 3 clusters= 0.506385941107445"
- > print(clusters[which.max(res[,2]),])

#### K=4

- [1] "max S for 4 clusters= 0.447485902399782"
- > print(clusters[which.max(res[,2]),])
- [1] 1 2 1 1 1 1 1 2 2 3 2 3 3 3 2 4 1 1 1 1 1 1 2 2 2 2 4 3 3 3 3 4 4 4 1 1 4 [36] 4 4 1

#### K=5

- [1] "max S for 5 clusters= 0.39902616181273"
- > print(clusters[which.max(res[,2]),])
- [1] 1 2 1 1 1 1 2 3 4 3 4 4 4 3 5 1 1 2 2 1 3 3 3 2 5 4 4 4 4 5 5 5 1 2 5 [36] 5 5 1

# Lampiran 5. (Lanjutan) **Data 2**K=3

[1] "max S for 3 clusters= 0.491430012258013"

#### K=4

[1] "max S for 4 clusters= 0.453603828225715"

> print(clusters[which.max(res[,2]),])
[1] 1 1 2 2 2 1 1 1 3 1 3 3 3 3 4 2 2 1 1 2 1 1 1 1 4 3 3 3 3 2 4 4 2 2 4
[36] 4 4 2

#### K=5

[1] "max S for 5 clusters= 0.430860000414149"

> print(clusters[which.max(res[,2]),])
[1] 1 1 2 2 2 1 1 3 4 1 4 4 4 4 5 2 2 1 1 2 3 3 3 1 5 4 4 4 4 2 5 5 2 2 5
[36] 5 5 2

Lampiran 5. (Lanjutan) **Data 3**K=3

[1] "max S for 3 clusters= 0.467607601210733"

#### K=4

[1] "max S for 4 clusters= 0.474614075642384"

#### K=5

[1] "max S for 5 clusters= 0.476373718721671"

> print(clusters[which.max(res[,2]),])
[1] 1 2 1 1 1 1 2 3 3 2 3 3 3 3 4 4 4 2 2 1 2 2 2 2 1 3 3 3 3 5 5 5 4 4 5
[36] 5 5 4

# Lampiran 5. (Lanjutan) **Data 4**K=3

[1] "max S for 3 clusters= 0.435117746440277"

#### K=4

[1] "max S for 4 clusters= 0.395242791352964"

#### K=5

[1] "max S for 5 clusters= 0.445598342860318"

Lampiran 6. Hasil Pengelompokan Tahun 2005, Tahun 2006, Tahun 2007 dan Tahun 2008 metode *K-means* pada Indeks Validitas *Silhouette* dengan *k*=3

Tahun 2005

Kelompok	Kabupaten/Kota	IHH	IP	PPP	IPM
4	Pacitan	75.5	71.35	54.21	67.02
	Ponorogo	72	67.04	56.4	65.15
	Trenggalek	75.67	75.44	54.97	68.69
	Blitar	74.83	73.47	57.07	68.46
	Kediri	72.5	75.08	56.11	67.9
	Malang	70.33	70.68	56.76	65.92
5	Banyuwangi	67.67	70.2	55.7	64.52
	Mojokerto	73.17	77.49	55.84	68.83
1	Jombang	73.33	77.13	56.63	69.03
1	Nganjuk	71	73.52	56.57	67.03
	Madiun	71.5	72.16	55.99	66.55
	Magetan	74.83	75.25	56.16	68.75
	Ngawi	72	67	54.91	64.64
	Bojonegoro	69.17	69.13	53.97	64.09
	Tuban	69.17	68.78	55.13	64.36
	Lamongan	70.67	70.92	56.61	66.06
	Kota Probolinggo	72.67	75.86	57.11	68.55
	Kota Pasuruan	68.33	82.47	56.07	68.96
	Tulungagung	75.67	75.1	58.01	69.59
	Sidoarjo	73.5	85.71	57.78	72.33
	Gresik	74.67	79.24	59.1	71
2	Kota Kediri	73.83	83.66	57.21	71.57
	Kota Blitar	76.67	84.76	58.13	73.19
47	Kota Malang	72.67	85.12	59.48	72.42
	Kota Mojokerto	76	84.74	58.93	73.22

Kelompok	Kabupaten/Kota	IHH	IP	PPP	IPM
4.73	Kota Madiun	75	85.75	60.01	73.59
2	Kota Surabaya	74.17	86.68	61.8	74.21
3:01	Kota Batu	71.83	79.25	57.45	69.51
TO	Lumajang	68.17	64.03	54.55	62.25
	Jember	61.17	65.95	53.02	60.05
	Bondowoso	60.83	60.82	53.16	58.27
	Situbondo	61.33	61.79	54.28	59.14
3	Probolinggo	58.33	61.46	54.23	58.01
	Pasuruan	62.17	72.24	54.86	63.09
5	Bangkalan	61.67	64.72	53.7	60.03
	Sampang	58	50.93	52.56	53.83
	Pamekasan	62	65.06	53.07	60.05
	Sumenep	64	64.88	54.86	61.24

# Tahun 2006

Kelompok	Kabupaten/Kota	IHH	IP	PPP	IPM
	Pacitan	75.75	73.2	53.7	67.55
	Ponorogo	72.58	67.13	56.56	65.43
	Trenggalek	76.08	76.28	54.4	68.92
	Tulungagung	75.75	74.94	56.13	68.94
	Kediri	72.58	75.66	54.76	67.67
1	Malang	70.5	73.52	55.23	66.41
	Lumajang	68.33	69.39	53.91	63.88
	Banyuwangi	68.25	71.41	55.3	64.99
	Mojokerto	73.67	75.91	56.01	68.53
	Jombang	73.83	75.45	55.44	68.24
	Nganjuk	71.08	73.76	54.49	66.44
	Madiun	71.83	71.04	54.92	65.93

Kelompok	Kabupaten/Kota	IHH	IP	PPP	IPM
435 6	Magetan	74.83	74.86	55.25	68.31
	Ngawi	72.58	63.7	54.5	63.59
	Bojonegoro	69.33	70	53.69	64.34
	Tuban	69.5	68.8	53.58	63.96
	Lamongan	70.83	71.42	55.7	65.99
	Kota Pasuruan	68.83	81.81	55.43	68.69
	Blitar	75.58	75.9	57	69.49
	Sidoarjo	73.75	85.69	56.92	72.12
2	Gresik	75.17	81.78	57.23	71.39
	Kota Kediri	(474)	82.85	56.3	71.05
	Kota Blitar	77.08	84.73	58.37	73.4
2	Kota Malang	73.33	87.76	61.08	74.06
	Kota Probolinggo	73.08	74.77	57	68.28
	Kota Mojokerto	76.08	85.5	57.98	73.19
	Kota Madiun	75.08	84.2	59.39	72.89
	Kota Surabaya	74.42	85.43	58.69	72.84
	Kota Batu	72.5	79.86	56.41	69.59
	Jember	61.67	68.44	52.16	60.75
	Bondowoso	61.48	58.54	52.7	57.57
	Situbondo	61.42	62.4	53.68	59.17
3	Probolinggo	58.67	61.71	53.96	58.11
	Pasuruan	62.25	72.01	54.69	62.98
	Bangkalan 💛	62	65.77	52.61	60.13
	Sampang	58.5	50.07	52.52	53.7
	Pamekasan	62.42	64.41	53.2	60.01
	Sumenep	64.42	61.91	55.15	60.49

Tahun 2007

Kelompok	Kabupaten/Kota	IHH	IP	PPP	IPM
Merry	Pacitan	76.31	76.2	53.81	68.77
ERSILL	Ponorogo	73.29	71.43	55.66	66.79
HITE.	Trenggalek	76.5	75.44	55.01	68.98
	Kediri	72.58	78.7	54.74	68.67
	Malang	70.46	74.58	55.16	66.73
	Banyuwangi	68.23	72.22	54.95	65.13
1	Jombang	74.68	78.73	56.57	70
	Nganjuk	70.99	75.32	55.34	67.22
	Madiun	71.66	72.49	55.03	66.39
	Magetan	74.88	74.77	55.84	68.5
	Ngawi	72.78	71.13	54.16	66.02
	Bojonegoro	69.49	71.04	54.92	65.15
	Tuban	69.83	69.99	53.6	64.47
	Lamongan	70.9	72.98	54.97	66.28
	Tulungagung	75.84	79.55	55.31	70.23
	Blitar	76.02	75.25	56.1	69.13
	Sidoarjo	73.9	86.07	56.18	72.05
	Mojokerto	73.78	80.37	56.05	70.07
	Gresik	75.59	79.86	54.89	70.11
	Kota Kediri	74.13	85.99	57.34	72.49
2	Kota Blitar	77.53	85.54	56.77	73.28
2	Kota Malang	74.23	87.21	58.34	73.26
	Kota Probolinggo	73.5	80.45	57.17	70.38
	Kota Pasuruan	68.96	84.07	56.15	69.73
	Kota Mojokerto	76.26	86.27	58.38	73.64
	Kota Madiun	75.35	88.46	59.28	74.36
TINUA !	Kota Surabaya	74.5	87.96	58.26	73.57
	Kota Batu	73.33	83.73	55.46	70.84

Kelompok	Kabupaten/Kota	IHH	IP	PPP	IPM
V P	Lumajang	68.5	67.98	52.1	62.86
411	Jember	62.42	68.85	52.46	61.24
Diff	Bondowoso	62.47	60.3	51.14	57.97
	Situbondo	61.3	66.79	53.37	60.48
3	Probolinggo	58.69	62.6	52.48	57.92
3	Pasuruan	62.33	73.11	51.88	62.44
	Bangkalan	62.33	63.65	52.71	59.56
	Sampang	59.45	49.91	52.79	54.05
	Pamekasan	62.83	65.27	53.23	60.44
	Sumenep 💢 🗴	65.49	64.53	54.61	61.54

# Tahun 2008

Kelompok	Kabupaten/Kota	IHH	IP	PPP	IPM
	Pacitan	77.35	76.26	54.96	69.52
	Ponorogo	74.33	71.47	58.71	68.17
	Kediri 🙆	73.24	77.92	56.67	69.28
	Malang	70.92	74.51	56.92	67.45
	Lumajang	69.32	71.11	55.02	65.15
	Banyuwangi	69.21	73.05	58.39	66.88
1	Nganjuk	71.27	76.72	58.12	68.7
	Madiun	72.55	74.59	59.58	68.91
	Ngawi	73.7	70.79	56.01	66.83
	Bojonegoro	70.19	70.48	57.95	66.21
	Tuban	70.47	70.55	57.89	66.3
	Lamongan	71.4	73.75	58.8	67.98
	Kota Batu	73.94	84	53.5	70.48

Kelompok	Kabupaten/Kota	IHH	IP	PPP	IPM
MASS E	Trenggalek	77.6	77.65	57.81	71.02
	Tulungagung	76.79	79.21	60.21	72.07
	Blitar	77.38	76.83	58.38	70.87
TO	Sidoarjo	74.58	86.27	59.06	73.3
	Mojokerto	74.49	79.68	58.53	70.9
	Jombang	75.74	78.84	59.49	71.36
	Magetan	75.1	77.04	59.06	70.4
2	Gresik	76.37	81.31	58.92	72.2
	Kota Kediri	74.71	86.8	58.66	73.39
	Kota Blitar	78.5	86.59	59.82	74.97
	Kota Malang	75.36	88.37	55.64	73.13
	Kota Probolinggo	74.23	79.59	58.52	70.87
	Kota Pasuruan	69.91	83.59	60.76	71.42
	Kota Mojokerto	77.22	86.33	61.18	74.91
	Kota Madiun	75.77	87.85	61.15	74.93
	Kota Surabaya	75.09	87.96	65.24	76.1
	Jember	63.32	68.87	54.8	62.33
	Bondowoso	63.63	61.58	52.15	59.12
	Situbondo	61.59	64.52	54.49	60.2
3	Probolinnggo	59.09	63.77	53.82	58.89
	Pasuruan	62.76	73.55	55.89	64.07
	Bangkalan	62.79	65.75	57.26	61.94
	Sampang	60.74	51.38	55.17	55.77
	Pamekasan	63.26	66.58	53.91	61.25
84	Sumenep	66.51	64.31	55.63	62.15